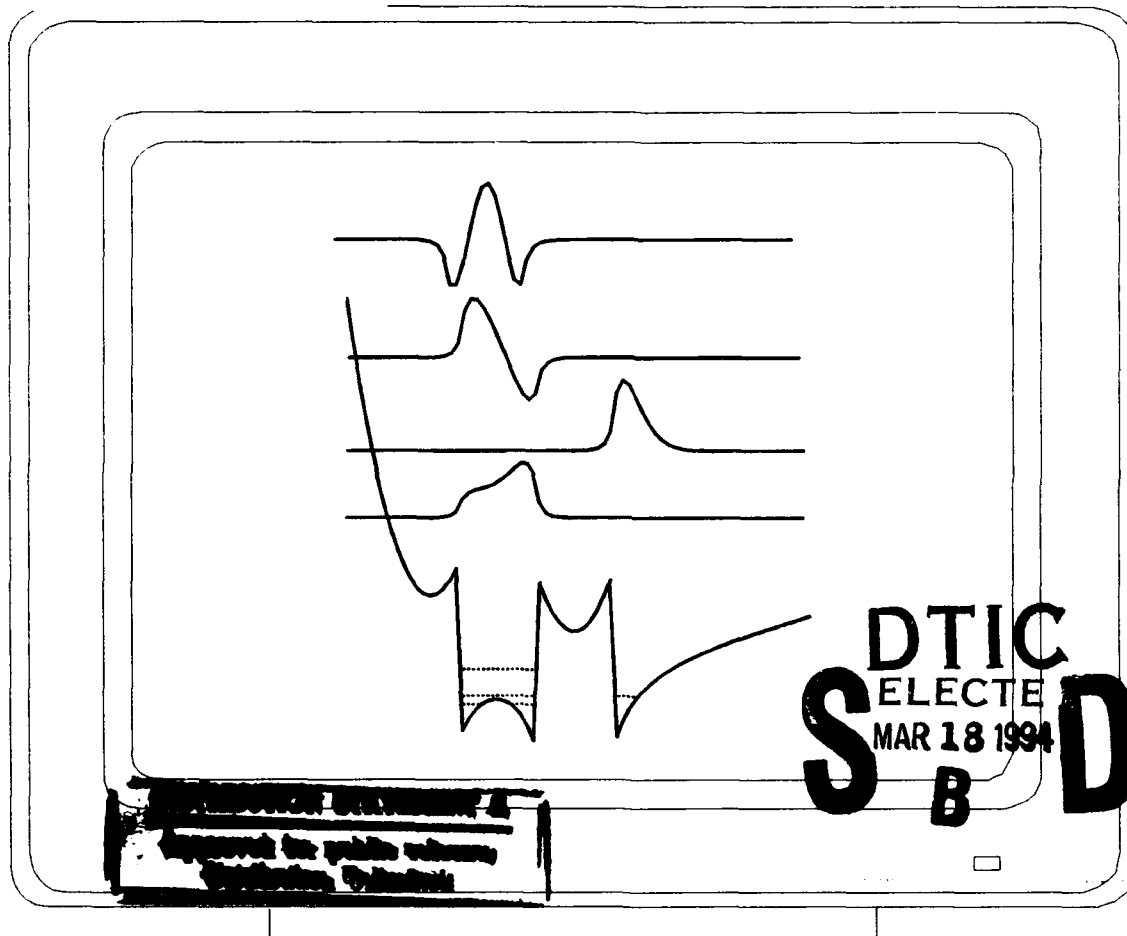


①

Proceedings of the International Workshop on Computational Electronics

AD-A277 020



DTIC
ELECTE
MAR 18 1994
S B D

University of Leeds
August 11-13th 1993

3286 94-08511

DTIC QUALITY INSPECTED

94 3 16 045

*Cover Illustration: Energy-band diagram and wavefunctions for a pHEMT structure.
R. Drury, Sponsored by UK SERC and M/A-COM (USA).*

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patent Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the authors, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency.

© University of Leeds and Authors Named in Each Contribution

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

Typesetting: Camera ready by contributors.
Printed by University of Leeds Printing Service

**Proceedings of the International
Workshop on Computational Electronics**

**Weetwood Hall
University of Leeds
England**

August 11th-13th 1993

Preface

The 1993 International Workshop on Computational Electronics took place at Leeds in England. The aim of the workshop is to allow researchers to focus on the computational aspects of solid-state device modelling. This encompasses new device and carrier transport models, numerical and computational techniques, computational implementation of transport theory (including quantum), developments in modelling electronic structures, new techniques in computer visualization of models. This year an underlying theme associated with compound semiconductor devices was adopted, but contributions on all aspects of solid-state modelling for all types of material were invited.

The workshop consists of several sessions, with invited papers, contributed papers and poster sessions, allowing opportunity for questions and technical interaction. Tutorial sessions are included at the start of the meeting to allow experts to address developments in computational numerical techniques.

This International Workshop on Computational Electronics follows the very successful series of workshops held in previous years in the United States of America. The earlier workshops were organised by the National Center for Computational Electronics (NCCE) and supported by the US National Science Foundation. This year we are grateful to the US Office of Naval Research, the US Army European Research Office and the Department of Electronic and Electrical Engineering at The University of Leeds for their support.

We are grateful to the members of the Programme Committee for reviewing the contributed papers and to the Advisory Committee for their helpful comments and suggestions. Over 60 contributed papers were submitted this year. The very high standard of these papers has led to a unique collection of contributions in this area. We are grateful to the invited speakers for providing keynote addresses.

Christopher M. Snowden
Michael J. Howes

University of Leeds, August 1993

Program Committee:

Co-Chairmen:

C.M. Snowden, *University of Leeds, UK*

M.J. Howes, *University of Leeds, UK*

Members:

R.W. Dutton, *Stanford University, USA*

D.K. Ferry, *Arizona State University, USA*

M. Fukuma, *NEC, Japan*

K. Hess, *University of Illinois, USA*

J. Jerome, *Northwestern University, USA*

M. Littlejohn, *North Carolina State University*

J. Meza, *Sandia National Laboratories, USA*

R.E. Miles, *University of Leeds, UK*

M. Pinto, *AT&T Bell Laboratories, USA*

U. Ravaioli, *University of Illinois, USA*

E. Sangiori, *University of Bologna, Italy*

P. Vogl, *Walter Schottky Institut, Germany*

Advisory Committee:

H. Bennett, *NIST*

F. Buot, *Naval Research Laboratory*

J. Cole, *Rensselaer Polytechnic Institute*

L. Cooper, *Office of Naval Research*

W.M. Coughran, *AT&T Bell Laboratories*

W. Fichtner, *ETH Zurich (Switzerland)*

M. Fischetti, *IBM Yorktown Heights*

W. Frensley, *University of Texas at Dallas*

S. Goodnick, *Oregon State University*

H. Grubin, *Scientific Research Associates*

C. Hamaguchi, *Osaka University (Japan)*

G.J. Iafrate, *Army Research Office*

T. Kerkhoven, *University of Illinois*

S. Laux, *IBM, Yorktown Heights*

T. McGill, *California Institute of Technology*

W. Porod, *University of Notre Dame*

C. Ringhofer, *Arizona State University*

D. Rose, *Duke University*

M. Schultz, *Yale University*

K. Smith, *AT&T Bell Laboratories*

T. Sugano, *Toyo University (Japan)*

A. Tasch, *University of Texas, Austin*

T. Thurgate, *Intel*

P. Yang, *Texas Instruments*

Accession For	
NTIS	GRA&I <input checked="" type="checkbox"/>
DTIC	IAS <input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
perform 50	
By	
Distribution	
Availability Codes	
Dist	Attn and/or Special
A-1	

K. Yokoyama., *NTT (Japan)*
Z. Yu, *Tsinghua University (China)*

Sponsors:

L.R. Cooper, Office of Naval Research
K. Steinbach, European Research Office, United States Army

Table of Contents

Tutorial Session

Numerical modelling of microwave devices (<i>Invited</i>) <i>Eric A B Cole, University of Leeds</i>	11
--	----

The classical and quantum hydrodynamic models (<i>Invited</i>) <i>Carl L Gardner, Duke University</i>	25
--	----

Conventional Simulation Techniques I

Technology computer-aided design (<i>Invited</i>) <i>S Selberhert, University of Vienna</i>	37
--	----

Application of quadrilateral finite elements for simulation of recess T-gate MESFETs and HEMTs <i>A Asenov, D Reid, J Barker, N Cameron and S Beaumont, University of Glasgow</i>	45
--	----

Numerical simulation of IGBTs at elevated temperatures <i>A Brown, A Asenov and J R Barker, University of Glasgow</i>	50
--	----

Deterministic BJT modeling by self-consistent solution to the Boltzmann, Poisson and Hole-continuity equations <i>Hongchin Lin, Neil Goldsman and I D Mayergoyz, University of Maryland</i>	55
--	----

Poster Session I

A self-consistent charge-control model for HEMTs incorporating deep level effects <i>Ranjit Singh and Christopher M Snowden, University of Leeds</i>	60
---	----

Transport coefficients for a GaAs hydrodynamic model extracted from inhomogeneous Monte Carlo calculations <i>Mei-Kei leong and Ting-wei Tang, University of Massachusetts</i>	65
---	----

Computation of the three-dimensional depletion approximation by Newton's method and multigrid <i>Frankline Bodine, Michael Holst and Thomas Kerkhoven, University of Illinois</i>	70
--	----

An accurate and efficient method to characterise heterostructure devices <i>Ali Abou-Elnour and Klaus Schuenemann, Technische Universität Hamburg</i>	75
--	----

New Simulation Approaches I

A new, easy-to-code, robust and stable approach to 2-D hydrodynamic submicron device modeling <i>Qi Lin, Neil Goldsman and Gwo-Chung Tai, University of Maryland</i>	80
---	----

An equivalent circuit model and distribution-function theory of the high frequency behaviour of quantum-based devices <i>F A Buot and K Rajagopal, Naval Research Laboratory</i>	85
Device simulation by means of a direct solution of the coupled Poisson/Boltzmann transport equations <i>C J Donnelly and C Lyden, University College, Cork</i>	91
Highly efficient simulation of HEMTs and MESFETs based on quantum mechanics <i>R K Veresegyhazy and C M Snowden, University of Leeds</i>	96
Special Evening Session (Invited)	
The future with computational electronics: a new golden age <i>J Barker, University of Glasgow</i>	101
New Simulation Approaches II	
Discretisation of the hydrodynamic transport equation using the control region approximation (<i>Invited</i>) <i>P Mawby and A McCowen, University College Swansea</i>	109
2-dimensional solution to the Boltzmann transport equation to arbitrarily high-order accuracy <i>Ken Hennacy, Neil Goldsman and I D Mayergoyz, University of Maryland</i>	118
Comparison of cellular automata and MINIMOS simulations of submicron MOSFETs <i>A Rein, G Zandler and P Lugli, Walter Schottky Institut</i>	123
Two-dimensional quantum modelling of heterojunction field effect transistors <i>R Drury, R E Miles and C M Snowden, University of Leeds</i>	128
Poster Session II	
Determination of diffusion coefficients and mobilities using the effective field concept <i>Kenneth Hennacy, Neil Goldsman and Issak Mayergoyz, University of Maryland</i>	133
Algorithms for the solution to the Helmholtz equation in the numerical simulation of semiconductor lasers <i>M Gault, P A Mawby and M S Towers, University College Swansea</i>	138
Visualisation techniques for semiconductor bandstructures <i>Alan Beck, P D Yoder and Umberto Ravaioli, University of Illinois</i>	146

Computation of the electron distribution function employing the theory of stochastic differential equations <i>Can E Korman, A Piazza and P Rugkwamsook, George Washington University and University of Maryland</i>	151
---	-----

Parallelization and Vectorization

PMC-3D: A parallelized 3D Monte Carlo simulator for electronic and electro-optic devices <i>S Pennathur, U A Ranawake and V K Tripathi, Oregon State University</i>	156
--	-----

Parallel simulation of semiconductor devices on MIMD machines <i>D Reid, A Asenov, J R Barker and S P Beaumont, University of Glasgow</i>	161
--	-----

Numerical simulation of a GaAs MESFET devices using parallel processing techniques with adaptive meshing and dynamic load balancing on a transputer network <i>C S Tsang-Ping, D M Barry and C M Snowden, University of Leeds</i>	166
--	-----

The implementation and speed-up of coloured SOR methods for solving the 3D Poisson equation on an array of transputers <i>A Brown, D Reid, A Asenov and J Barker, University of Glasgow</i>	171
--	-----

Electronic Structure Calculations and Applications

First principles calculation of electron-phonon deformation potentials in silicon <i>P D Yoder, University of Illinois</i>	176
---	-----

A CPA calculation for disorder-induced intervalley scattering in AlGaAs <i>Selim E Günçer and D K Ferry, Arizona State University</i>	181
--	-----

Steady-state electron transport in silicon dioxide employing different electronic band-structures <i>M Hackel, H Kosina and S Selberherr, T U Vienna</i>	186
---	-----

Monte Carlo Simulation I

New developments in Monte Carlo device simulation (<i>Invited</i>) <i>U Ravaoli, Beckman Institute</i>	192
---	-----

Poster Session III

Electron-hole scattering effects in silicon <i>Jim Dewey and Mohamed A Osman, Washington State University</i>	200
--	-----

Hole transport in the warped band model of GaAs <i>N Nintunze and M A Osman, Washington State University</i>	205
Ensemble Monte Carlo simulation of the hot electron transport in the heterojunction bipolar transistors <i>G Khrenov, V Ryzhii and S Kartashov, Russian Academy of Sciences</i>	209
Scaled ensemble Monte Carlo <i>A M Krizan and R P Joshi, University at Buffalo and Old Dominion University</i>	214
Acoustic phonon controlled transport in low dimensional structures <i>V Mitin, R Mickevicius and N Bannov, Wayne State University</i>	219
Monte Carlo Simulation II	
Highly efficient full-band Monte Carlo simulation (<i>Invited</i>) <i>K Smith, AT & T</i>	224
Ionized impurity scattering model for Monte Carlo calculations <i>J Dewey and M A Osman, Washington State University</i>	231
A Monte Carlo method for the study of non-equilibrium phenomena in low dimensional semiconductor structures <i>L Rota, Fausto Rossi, P Lugli and E Monlinari, University of Oxford, Philipps-Universität Marburg, Università di Roma and Università di Modena</i>	236
Monte Carlo modelling of sub-micron delta-doped MOSFETs <i>R W Kelsall and A G O'Neill, University of Newcastle-upon-Tyne</i>	241
Quantum Simulation Techniques I	
Techniques for modelling ultra-small quantum devices (<i>Invited</i>) <i>D K Ferry, Arizona State University</i>	247
Boundary conditions for quantum devices with exposed surfaces <i>Minhan Chen and Wolfgang Porod, University of Notre Dame</i>	255
Queueing-theoretic simulation of single-electronic metal-semiconductor devices and systems <i>Sharif Babiker, John R Barker and Asen Asenov, University of Glasgow</i>	260
Simulation of Si/SiGe modulation-doped FET using quantum hydrodynamic equations <i>J R Zhou, T Yamada, H Miyata and D K Ferry, Arizona State University</i>	265

Poster Session IV

Parallel algorithms for the simulation and visualisation of interacting few-electron quantum transport in laterally patterned low-dimensional semiconductors 270

J Cluckie and J R Barker, University of Glasgow

System simulation tools for single-electronic devices 275

Scott Roy, John R Barker and Asen Asenov, University of Glasgow

Adaptive potential in alternating-direction implicit relaxation 280

Nam Young Kim and A M Krimin, State University of New York at Buffalo

Numerical study of 2-D quantum dots 285

M Macucci and Karl Hess, Università di Pisa and University of Illinois

Numerical methods for Bloch-Poisson type equations 291

Norbert J Mauser, Christian Ringhofer and Gabriele Ulich, Technische Universität Berlin and Arizona State University

Quantum Simulation Techniques II

Transport simulations for quantum well heterostructures and lasers 296

(Invited)

K Hess, University of Illinois

Quantum simulation of several-particle systems 303

Craig Lent, Wolfgang Porod and P Douglas Tougaw, University of Notre Dame

Full-scale self consistent simulation of quantum devices 308

Dejan Jovanovic and Jean-Pierre Leburton, University of Illinois

Planar supercell simulations of 3D quantum transport in semiconductor nanostructures 313

D Z-Y Ting, S K Kirby and T C McGill, California Institute of Technology

Quantum Simulation Techniques III

Simulation of coherent quantum transport in a magnetic field: recovery of conductance quantization 318

Manhua Leng and Craig Lent, University of Notre Dame

Numerical method to obtain the two-dimensional electronic states for open boundary scattering problems 323

Henry Harbury and Wolfgang Porod, University of Notre Dame

A numerical study of resonance/antiresonance line shape for transmission in quantum waveguides with resonantly-coupled cavities <i>Zhi-an Shao, Wolfgang Porod and Craig Lent, University of Notre Dame</i>	328
Effects of optical and acoustic phonon scattering on hot electron transport in quantum wires <i>N Mori, H Momose, S Miyahara and C Hamaguchi, Osaka University</i>	333

Numerical Modelling of Microwave Devices

Eric A B Cole

Department of Applied Mathematical Studies,
University of Leeds,
Leeds LS2 9JT, UK

Abstract

Various numerical routines are discussed and applied to the numerical simulation of a planar sub-micron gate length GaAs MESFET. The model contains energy transport and degenerate statistics. A generalisation of the Scharfetter-Gummel method is given which enables both the electron current density and energy flux to be easily coded. Degenerate statistics are included for the case in which the relaxation time is proportional to E^r . Steady state results are presented - these are arrived at both by iterating the transient solution and the direct method based on a modified Newton method.

I Introduction

In this paper we look at the finite difference approach to the modelling of a two-dimensional MESFET. Other methods - the Finite Element method [1,2], the Boundary Element method and the Multigrid method [3] can be found elsewhere. To illustrate the general theory developed in Section IV, we will consider the simulation of a planar sub-micron gate length GaAs MESFET whose cross section in the x-y plane is shown in figure 1, with the ends of the source, gate and drain at $x=s_0, s_1, g_0, g_1, d_0$ and d_1 . The electron density n , electron temperature T_e and potential ψ will all be functions of x, y and time t . The equations are:

(i) the Poisson equation

$$\nabla^2 \psi = \frac{q}{\epsilon} (N_d - n) \quad (1)$$

where ϵ is the product of the permittivity of the vacuum and the relative permittivity of the semiconductor. This equation is to be solved with the boundary conditions $\psi=V_s$ on the source, $\psi=V_d$ on the drain, $\psi=V_g+\phi_b$ on the gate where ϕ_b is the built-in potential, and $\partial\psi/\partial n = 0$ at other parts of the

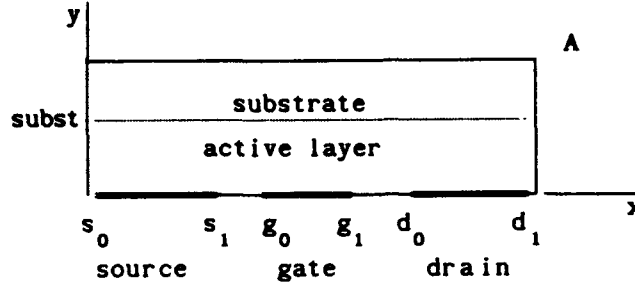


Fig. 1. Cross section of the MESFET.

boundary (In this paper, \underline{n} will represent the unit vector normal to the surface). The electric field is

$$\underline{E} = -\nabla\psi ; \quad (2)$$

(...), the current continuity equation (neglecting recombination)

$$\frac{\partial n}{\partial t} = \frac{1}{q} \nabla \cdot \underline{J} \quad (3)$$

where the current density \underline{J} will have the form

$$\underline{J} = \alpha_c n(\nabla\psi) + \beta_c T_e (\nabla n) + \gamma_c n(\nabla T_e) \quad (4)$$

for certain coefficients α_c , β_c and γ_c . The boundary conditions taken here are $n=2.5N_d$ on the source and drain, $n=0$ on the gate, and $\partial n/\partial \underline{n}=0$ elsewhere;

(iii) the energy transport equation

$$\frac{\partial W}{\partial t} = \underline{J} \cdot \underline{E} - \frac{(W-W_0)}{\tau(\xi)} - \nabla \cdot \underline{s} \quad (5)$$

where ξ is the (position and time-dependent) average electron energy, $W=n\xi$, $\tau(\xi)$ is the energy-dependent relaxation time, and \underline{s} is the energy flux

$$\underline{s} = \alpha_E W(\nabla\psi) + \beta_E T_e (\nabla W) + \gamma_E W(\nabla T_e) \quad (6)$$

for certain coefficients α_E , β_E and γ_E . The forms of all the coefficients will be discussed more fully in section V when particular statistics are introduced.

Having obtained a set of modelling equations, they should always be scaled to ensure a good numerical range for the variables [1,4].

II Finite Differences

We will deal only with the special case of a non-uniform rectangular mesh. More general techniques including, for example, mesh refinement [1,5] and box generation [6] are described elsewhere. Consider the two-dimensional rectangular mesh shown in figure 2. Mesh points will be labelled $0,1,2,\dots,M$ and $0,1,2,\dots,N$ in the x and y directions respectively. The general mesh point will have coordinate (x_i, y_j) . The variable mesh spacings will be $h_i = x_{i+1} - x_i$ ($i=0,\dots,M-1$) and $k_j = y_{j+1} - y_j$ ($j=1,\dots,N-1$). The case of uniform mesh is given by $h_i = h = \text{const.}$ and $k_j = k = \text{const.}$ The value $f(x_i, y_j)$ of any function f will be denoted shortly by $f_{i,j}$, while its value at the half-points $(x_i + h_i/2, y_j)$, $(x_i, y_j + k_j/2)$ and $(x_i + h_i/2, y_j + k_j/2)$ will be denoted by $f_{i+1/2,j}$, $f_{i,j+1/2}$ and $f_{i+1/2,j+1/2}$ respectively.

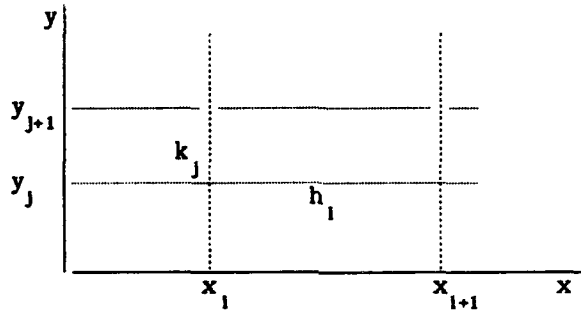


Fig. 2. The mesh.

Standard formulae exist for first and second order derivatives [7]. The important ones for this simulation are listed below (resorting to the one-dimensional case temporarily to ease the notation). The half-point formula for the first derivative of f is given by

$$f'_1 = \frac{f_{i+1/2} - f_{i-1/2}}{(h_i + h_{i-1})/2} \quad (7)$$

where the next term $-(h_i - h_{i-1})f''_1/4$ has been neglected. This is useful for evaluating divergences. The second order derivative at x_i can be written

$$f''_1 = \frac{f'_{i+1/2} - f'_{i-1/2}}{(h_i + h_{i-1})/2} = \frac{(f_{i+1} - f_i)/h_i - (f_i - f_{i-1})/h_{i-1}}{(h_i + h_{i-1})/2}$$

and the neglected term here is $(h_i - h_{i-1})f'''_1/3$.

We often need an expression for the normal derivative at a boundary, in

which case it is very often zero. False nodes can be introduced outside the boundary but this increases storage requirements. Alternatively, this derivative may be expressed at $x=0$ in terms of values at internal nodes by

$$f'_0 = - \frac{h_1(h_1+2h_0)f_0 - (h_0+h_1)^2 f_1 + h_0^2 f_2}{h_0 h_1 (h_0+h_1)} \quad (8)$$

with a similar expression at $x=x_M$.

Returning to the case of functions of two variables, the Poisson equation $\nabla^2 \psi = g$ can be discretised in the form

$$\frac{\frac{\psi_{1+1,j} - \psi_{1,j}}{h_1} - \frac{\psi_{1,j} - \psi_{1-1,j}}{h_{1-1}}}{(h_1+h_{1-1})/2} + \frac{\frac{\psi_{1,j+1} - \psi_{1,j}}{k_j} - \frac{\psi_{1,j} - \psi_{1,j-1}}{k_{j-1}}}{(k_j+k_{j-1})/2} - g_{1,j} = 0. \quad (9)$$

In the case of uniform mesh this reduces to the standard 5-point formula

$$p^2 \psi_{1-1,j} + p^2 \psi_{1+1,j} + \psi_{1,j-1} + \psi_{1,j+1} - 2(1+p^2) \psi_{1,j} = p^2 h^2 g_{1,j} \quad (10)$$

where $p=k/h$. This equation has an error of order h^4 , while the more accurate 9-point formula with error of order h^6 is [8]

$$\begin{aligned} & (10p^2-2)(\psi_{1-1,j} + \psi_{1+1,j}) + (10-2p^2)(\psi_{1,j-1} + \psi_{1,j+1}) \\ & + (1+p^2)(\psi_{1-1,j-1} + \psi_{1+1,j-1} + \psi_{1-1,j+1} + \psi_{1+1,j+1} - 20\psi_{1,j}) \\ & = p^2 h^2 (g_{1-1,j} + g_{1+1,j} + g_{1,j-1} + g_{1,j+1} + 8g_{1,j}) \end{aligned} \quad (11)$$

III Solution of Simultaneous Equations

The discretisation of the differential equations of the model generally gives rise to the problem of solving a set of $M+1$ analytical equations of the form

$$f_i(X_0, X_1, \dots, X_M) = 0 \quad (i=0, 1, \dots, M), \quad (12)$$

where the $M+1$ values X_i will represent typical physical quantities at each point i . If there are ν physical quantities described at each point, then the top index M is replaced by $\nu(M+1)-1$. The method of solution will depend on whether all the equations are linear or at least one equation is nonlinear.

A. Linear Equations

When equations (12) are linear they take the form

$$AX = b \quad (13)$$

where A is an $(M+1) \times (M+1)$ matrix and b is an $(M+1)$ component vector, both with elements which may depend on X . These can be solved by direct methods or by indirect methods using iteration.

(a) Direct Methods. Very often the matrix is tridiagonal and a simple solution is available. Suppose that the equations have the form

$$\alpha_i X_{i-1} + \beta_i X_i + \gamma_i X_{i+1} = \delta_i \quad (i=0,1,\dots,M)$$

and the terms involving α_0 and γ_M do not appear. The standard method of eliminating down from the first equation and solving the resulting final two equations for X_{M-1} and X_M gives the solution

$$\begin{aligned} X_M &= \left[(\beta'_{M-1})^{-1} \gamma_{M-1} - (\alpha_M)^{-1} \beta_M \right]^{-1} \left[(\beta'_{M-1})^{-1} \delta'_{M-1} - (\alpha_M)^{-1} \delta_M \right] \\ X_i &= (\beta'_i)^{-1} \left[\delta'_i - \gamma_i X_{i+1} \right] \quad (i=M-1, \dots, 0) \end{aligned} \quad (14)$$

where the β'_i and δ'_i are generated by

$$\left. \begin{aligned} \beta'_0 &= \beta_0, \quad \delta'_0 = \delta_0 \\ \beta'_i &= \beta_i - \alpha_i (\beta'_{i-1})^{-1} \gamma_{i-1} \\ \delta'_i &= \delta_i - \alpha_i (\beta'_{i-1})^{-1} \delta'_{i-1} \end{aligned} \right\} \quad (i=1, \dots, M-1).$$

Efficient coding of this routine is available [9,10]. We can usually get away without pivoting in this case because the simulation will probably not cause problems with the sizes of the elements of the tridiagonal system. In one-dimensional simulations which involve ν physical variables at each node (for example, $\nu=3$ when the variables are n , p and ψ), the equations at each node can be grouped so that equations (14) still hold. In that case, however, the X_i , δ_i and δ'_i are ν -component vectors and the quantities α_i , β_i , γ_i and β'_i are $\nu \times \nu$ matrices. The multiplications in equations (14) are in the correct order for matrix multiplication [11].

When A is not simply tridiagonal, LU decomposition with partial pivoting

and scaling to control the growth of rounding errors may be used [10,12,13]. Rounding errors in direct methods can often be eliminated by solving the equation $A\tilde{x} = b - A\tilde{x}_c$ for the correction \tilde{e} to the computed solution \tilde{x}_c .

(b) Relaxation Methods. Consider the iterative process

$$\tilde{x}^{k+1} = B\tilde{x}^k + \tilde{c} \quad (15)$$

where the iteration matrix B is an $(M+1) \times (M+1)$ matrix with constant coefficients. If the solution converges to \tilde{x} then equation (15) is equivalent to the equation $(I-B)\tilde{x} = \tilde{c}$. Convergence may be slow if the spectral radius of B is close to unity, so instead take a modified iteration process

$$\tilde{x}^{k+1} = \tilde{x}^k + w \left(\tilde{c} - (I-B)\tilde{x}^k \right) = B'\tilde{x}^k + w\tilde{c} \quad (16)$$

and adjust the value of the parameter w so that the spectral radius of B' is as small as possible. This is the basis of the relaxation method.

In our original problem given by equation (13), we may write $A = L_0 + D + U_0$ where D is diagonal and L_0 and U_0 are lower and upper triangular with zeros on the diagonals. This decomposition suggests the Jacobi iteration

$$D\tilde{x}^{k+1} = \tilde{b} - (L_0 + U_0)\tilde{x}^k$$

with iteration matrix $B = -D^{-1}(L_0 + U_0)$, and the Gauss-Seidel iteration

$$(L_0 + D)\tilde{x}^{k+1} = -U_0\tilde{x}^k + \tilde{b}$$

with iteration matrix $B = -(L_0 + D)^{-1}U_0$. It can be shown [9,14,15] that the scheme (16) is convergent only for $0 < w < 2$. If $0 < w < 1$ then we have Successive Under-Relaxation (SUR) while if $1 < w < 2$ we have Successive Over-Relaxation (SOR). Further, the Jacobi spectral radius for equation (10) is

$$\rho_J = \frac{p^2 \cos \frac{\pi}{M+1} + \cos \frac{\pi}{N+1}}{1 + p^2}$$

with optimal choice

$$w = \frac{2}{1 + (1 - \rho_J^2)^{1/2}}$$

This is not the exact value for the equation (11) but it very often works well enough. See also [16]. The value of w may also be changed as the iteration progresses [17], for example, Chebyshev Acceleration.

B. Non-linear Equations

The standard method here is the Newton-Raphson method. If \underline{X}_c and $\underline{X} = \underline{X}_c + \delta \underline{X}$ are the computed and true solution, expansion of (12) to first order gives

$$J \delta \underline{X} = -f(\underline{X}_c) \quad \text{where} \quad J_{ij} = \left. \frac{\partial f_i}{\partial X_j} \right|_c$$

is the Jacobian. This is a set of linear equations which can be solved (among other methods) using the iteration scheme

$$J^k \delta \underline{X}^k = -f(\underline{X}^k), \quad \underline{X}^{k+1} = \underline{X}^k + \alpha \delta \underline{X}^k$$

where α is taken such that $0 < \alpha \leq 1$ to avoid overshoot. It is usually necessary to take α small in the early stages of the iteration and then to steadily increase it as the iteration progresses.

An alternative Newton method has been found to make iterations go more smoothly. Instead of solving for $\delta \underline{X}^k$ we solve directly for \underline{X}^{k+1} :

$$J \underline{X}^{k+1} = J(\underline{X}^k + \alpha \delta \underline{X}^k) = J \underline{X}^k - \alpha f^k. \quad (17)$$

The advantages here are that we can usually take larger values of α and that fewer initial guesses need to be made. When $J_{ii} \neq 0$ ($i=0, \dots, M$) then (17) becomes

$$X_1^{k+1} = J_{11}^{-1} \left((-\alpha f + J \underline{X}^k)_1 - \sum_{j=1}^M J_{1j} X_j^{k+1} \right)$$

with corresponding iteration process

$$X_1^{k+1} = (1-w) X_1^{k+1} + w J_{11}^{-1} \left((-\alpha f + J \underline{X}^k)_1 - \sum_{j=1}^M J_{1j} X_j^{k+1} \right) \quad (18)$$

where $f = f(\underline{X}^k)$ and $J = J(\underline{X}^k)$. If \underline{X} represents a natural grouping of ν quantities at each point (that is, X_1 is a ν -component vector) then equation (18) still holds with J_{11} being a $\nu \times \nu$ nonsingular matrix, and w as a $\nu \times \nu$ diagonal matrix with separate relaxation factors down the diagonal.

Evaluation of the Jacobian can be the most time-consuming part of the operation. It is possible to differentiate the functions f_i numerically using

$$\frac{\partial f_i}{\partial X_j} \approx \frac{f(X_0, \dots, X_j + \Delta X_j, \dots, X_M) - f(X_0, \dots, X_j, \dots, X_M)}{\Delta X_j}$$

where the ΔX_j are suitable increments. If the ΔX_j are too small then roundoff

errors can swamp the calculation, while convergence will be linear if they are too large [18]. Methods could be used in which, say, standard functions are differentiated explicitly while, for example, mobility curves fed in from other simulations could be differentiated numerically. Also using Broyden's rank one correction it is not necessary to evaluate J at each iteration [19].

IV Discretising the Current Continuity and Energy Equations

An explicit time discretisation of the current continuity equation will only work if the timestep Δt is excessively small [20], while a fully implicit Crank-Nicholson scheme [12] is difficult to solve. The linearised semi-implicit scheme [21] is more satisfactory unless Δt is taken too large. This limitation can be avoided if we use, in the case of constant T_0 , the Scharfetter-Gummel method in which we take an exponential variation in the carrier deviations between nodes [22,23].

The following is a generalisation of the Scharfetter-Gummel method when T_0 is not constant throughout the device. Equations (4) and (6) can be written in general form

$$\underline{V} = \alpha(\nabla\psi)\theta + \beta T_0(\nabla\theta) + \gamma\theta(\nabla T_0) \quad (19)$$

where $(\underline{V}=\underline{J}, \theta=n)$ for the current density and $(\underline{V}=\underline{S}, \theta=W)$ for the energy flux. For numerical purposes only we now make the assumptions that, in the interval $(i,j) \rightarrow (i+1,j)$, the quantities $\nabla\psi$, ∇T_0 , β , α/β and γ/β are constants. The x-component of equation (19) becomes

$$\frac{\partial\theta}{\partial x} + \left[\frac{\gamma}{\beta} \frac{1}{T_0} \frac{\partial T_0}{\partial x} + \frac{\alpha}{\beta} \frac{1}{T_0} \frac{\partial\psi}{\partial x} \right] \theta = \frac{1}{T_0} \left(\frac{V_x}{\beta} \right)_{i+1/2,j}$$

which has integrating factor

$$\exp \left[\int \left(\frac{\gamma}{\beta} \frac{1}{T_0} \frac{\partial T_0}{\partial x} + \frac{\alpha}{\beta} \frac{1}{T_0} \frac{\partial\psi}{\partial x} \right) dx \right] = T_0^r \quad (20)$$

where $r = \frac{\gamma}{\beta} + \frac{\alpha}{\beta} \frac{\partial\psi}{\partial x} \left(\frac{\partial T_0}{\partial x} \right)^{-1}$ and quantities γ/β etc are evaluated at $(i+1/2,j)$. The resulting equation is

$$\frac{\partial}{\partial x} (\theta T_0^r) = T_0^{r-1} \left(\frac{V_x}{\beta} \right)_{i+1/2,j}$$

which can be integrated between $x=x_i$ and $x=x_{i+1}$ and re-arranged to give

$$(V_x)_{1+1/2,j} = \frac{\beta_{1+1/2,j}}{h_1} \left[C_{\gamma/\beta} \left(-\frac{\alpha}{\beta} (\psi_{1+1,j} - \psi_{1,j}), T_{1+1,j}, T_{1,j} \right) \theta_{1+1,j} - C_{\gamma/\beta} \left(\frac{\alpha}{\beta} (\psi_{1+1,j} - \psi_{1,j}), T_{1,j}, T_{1+1,j} \right) \theta_{1,j} \right] \quad (21)$$

with a similar expression for $(V_y)_{1,j+1/2}$. Here, C is the function defined by

$$C_a(x, y, z) = p(y, z) B \left(\frac{x - a(y - z)}{p(y, z)} \right) \quad (y \neq z) \quad (22)$$

where $p(y, z) = \frac{(z - y)}{\ln(z/y)}$ and $B(t) = \frac{t}{e^t - 1}$ is the Bernoulli function.

One important property of this C -function, namely

$$\lim_{T \rightarrow T_0} [t C_a(x, T, T)] = T_0 B(x/T_0)$$

is useful for applying to models in which the electron temperature is a constant T_0 , and the original Scharfetter-Gummel expression is then obtained.

It is necessary to avoid overflows and underflows when evaluating the C -functions. The usual method [1] is to make a piecewise machine-dependent approximation for $B(t)$. The derivative of the Bernoulli function can be approximated by

$$B'(t) = \begin{cases} (t-2)/4 & \text{for } t_2 \leq t \leq t_3 \\ \frac{B(t)}{t} (1 - B(t) - t) & \text{otherwise.} \end{cases}$$

V Evaluation of the Coefficients

The Fermi integral $F_r(\eta)$ is defined as

$$F_r(\eta) = \Gamma(r+1)^{-1} \int_0^\infty y^r (1 + e^{y-\eta})^{-1} dy \quad \text{where} \quad \Gamma(r+1) = \int_0^\infty y^r e^{-y} dy$$

This function has been extensively approximated and tabulated [24]. Write k_B as Boltzmann's constant, h_p as Planck's constant, μ as the mobility, m_e as the effective electron mass, $\eta = (E_F - E_c)/(k_B T_e)$, and take the energy dependence of the relaxation time as $\tau \propto (E - E_c)^r$. The first few moments of the Boltzmann transport equation then give [25,26]

$$n = AT_{\bullet}^{3/2} F_{1/2}(\eta) \quad \text{where} \quad A = 2(2\pi k_B m_{\bullet})^{3/2} / h_p^3,$$

$$W = \frac{3}{2} k_B T_{\bullet} n F_{3/2}(\eta) / F_{1/2}(\eta),$$

$$\tilde{J} = \mu n (-q \nabla \psi + k_B T_{\bullet} \nabla \eta) + I_2 (\nabla T_{\bullet}) / T_{\bullet},$$

$$q \tilde{s} = -I_2 (-q \nabla \psi + k_B T_{\bullet} \nabla \eta) - I_3 (\nabla T_{\bullet}) / T_{\bullet},$$

where

$$I_2 = \mu n (r + \frac{5}{2}) k_B T_{\bullet} F_{r+3/2} / F_{r+1/2}$$

$$I_3 = I_2 (r + \frac{7}{2}) k_B T_{\bullet} F_{r+5/2} / F_{r+3/2}.$$

After extensive manipulation and using the constant effective mass approximation, it is found that \tilde{J} and \tilde{s} have the forms (4) and (6) where

$$\alpha_c = -q\mu$$

$$\beta_c = k_B \mu F_{1/2} / F_{-1/2}$$

$$\gamma_c = k_B \mu \left((r + \frac{5}{2}) F_{r+3/2} / F_{r+1/2} - \frac{3}{2} F_{1/2} / F_{-1/2} \right)$$

$$\alpha_E = \frac{2}{3} \mu (r + \frac{5}{2}) F_{r+3/2} F_{1/2} / (F_{r+1/2} F_{3/2})$$

$$\beta_E = -\frac{2}{3} k_B \mu (r + \frac{5}{2}) F_{r+3/2} / (q F_{r+1/2})$$

$$\gamma_E = \frac{2}{3} k_B \mu (r + \frac{5}{2}) \left(\frac{5}{2} F_{r+3/2} / F_{r+1/2} - (r + \frac{7}{2}) F_{r+5/2} / F_{1/2} / (F_{r+1/2} F_{3/2}) \right) / q$$

and $F_1 \equiv F_1(\eta)$ throughout.

At certain stages of the numerical implementation it is necessary to find η and T_{\bullet} at each point given the values of n and W . From the above, it may be shown that

$$F_{1/2}(\eta)^5 / F_{3/2}(\eta)^3 = 54 h_p^6 (8\pi m_{\bullet})^{-3} n^5 / W^3 \quad (23)$$

and this may be inverted to give η and hence T_{\bullet} from

$$T_{\bullet} = \frac{2}{3} W F_{1/2}(\eta) / [k_B n F_{3/2}(\eta)].$$

Conversely, to find η and W given n and T_{\bullet} , η may be calculated at each point by inverting the equation

$$F_{1/2}(\eta) = n/(AT_0^{3/2}) \quad (24)$$

and calculating W from

$$W = \frac{3}{2} k_B T_0 n F_{3/2}(\eta) / F_{1/2}(\eta).$$

Both these inversions must be done at each grid point. Once the functions $F_{1/2}$ and $F_{3/2}$ have been programmed, inversion is straightforward since the left hand sides of equations (23) and (24) are strictly increasing functions of η .

VI Implementation for the MESFET

We now apply the preceeding results to the simulation of the MESFET model outlined in section I. Referring to figure 1, the dimensions taken were $s_0=0.0\mu\text{m}$, $s_1=0.2\mu\text{m}$, $g_0=0.6\mu\text{m}$, $g_1=1.1\mu\text{m}$, $d_0=1.6\mu\text{m}$ and $d_1=1.8\mu\text{m}$. The total thickness including substrate was $0.45\mu\text{m}$. An abrupt junction at $y=0.35\mu\text{m}$ was taken with

$$N_d = \begin{cases} 10^{23} \text{ m}^{-3} & \text{for } 0\mu\text{m} \leq y \leq 0.35\mu\text{m} \\ 10^{19} \text{ m}^{-3} & 0.35\mu\text{m} < y \leq 0.45\mu\text{m}. \end{cases}$$

Monte Carlo simulations and experimental data on the steady state transport characteristics provide curves of ξ and τ in terms of the static electric field E_{ss} which is used as an intermediate parameter. This enables τ to be found in terms of ξ . The mobility is given by

$$\mu = \frac{300\mu_0}{T_0} \left[\frac{1 + \frac{8.5 \times 10^4 E^3}{\mu_0 (1 - 5.3 \times 10^{-4} T_0) E_0^4}}{1 + (E/E_0)^4} \right] \quad \text{where} \quad \mu_0 = \frac{0.8}{1 + (N_d/10^{23})^{1/2}}$$

and $E=4 \times 10^5 \text{ Vm}^{-1}$. All boundary conditions $\partial/\partial n=0$ were implemented using equation (8) and its equivalents with all derivatives zero. The quantities \underline{J} and \underline{s} were coded at the half points using appropriate values in expression (21). Three solutions were performed: the transient solution, the steady state reached by iterating the transient solution, and the direct solution of the steady state.

(a) The transient solution. A timestep $\Delta t=10^{-14} \text{ s}$ was used. The Poisson equation (11) was used to extract an expression $A_{i,j}$ for $\psi_{i,j}$ in terms of the other quantities, and was solved by iterating the equation

$$\psi_{i,j} = (1-w_{psi})\psi_{i,j} + w_{psi}A_{i,j}$$

The current continuity and energy equations (3) and (5) were similarly solved by coding the C functions with the appropriate quantities $\alpha_c, \dots, \gamma_E$. It was found that the separate relaxation factors $w_{psi}=1.4$, $w_n=0.8$ and $w_T=0.5$ gave the fastest iteration. At each timestep, the three equations were iterated separately inside an overall iteration. A mixture of convergence tests was used. In the case of the ψ iteration, we required convergence at each point, that is, we required that $|\psi_{i,j} - \psi_{i,j}|$ had to be less than some prescribed value at every point, where ψ_1 was the value at the previous iteration. The same test was applied to the electron temperature T_e . In the case of the $n_{i,j}$ we applied a weaker test by requiring that only the average relative difference $\sum |n_{i,j} - n_{i,j}| / n_{i,j}$ be smaller than some prescribed value. It was found that this average condition gave a smoother time plot of the total current

$$J_{tot} = \int_0^{0.45} \left(J_x + \epsilon \partial E_x / \partial t \right) dy$$

(b) Iteration to the steady state. The method of (a) was used but with no iteration of the equations at each timestep. A total of 2000 timesteps was used, giving a device time of 20ps. Results for ψ , n and T_e are shown in figure 3 (viewed from corner A in figure 1) for the case $r=-1$.

(c) Direct method. A natural grouping $X_{i,j} = (\psi_{i,j}, n_{i,j}, T_{e,i,j})^T$ exists at each gridpoint (i,j) . Put all $\partial/\partial t$ terms zero in equations (1), (3) and (5) and write the equations generally as $G_1=0$, $G_2=0$ and $G_3=0$. The Newton method applied at each point gives

$$A\delta X_{i-1,j} + B\delta X_{i+1,j} + C\delta X_{i,j} + D\delta X_{i,j-1} + E\delta X_{i,j+1} = -G$$

where $A_{mi} = \partial G_m / \partial \psi_{i-1,j}$ etc are 3×3 matrices. Writing $y_{i,j} = X_{i,j} + \alpha \delta X_{i,j}$, the modified Newton method (18) becomes

$$y_{i,j}^{k+1} = (I-w)y_{i,j}^{k+1} + wC^{-1} \left(-\alpha G + AX_{i-1,j}^k + BX_{i+1,j}^k + CX_{i,j}^k + DX_{i,j-1}^k + EX_{i,j+1}^k \right. \\ \left. - Ay_{i-1,j}^{k+1} - By_{i+1,j}^{k+1} - Dy_{i,j-1}^{k+1} - Ey_{i,j+1}^{k+1} \right)$$

where w is the diagonal matrix with nonzero elements w_{psi} , w_n and w_T . This routine is fragile in the early stages, and for the first few Newton iterations it is necessary to take α very small (typically 0.01), $n_{i,j}$ to be kept positive and $T_{e,i,j}$ to be not less than T_0 .

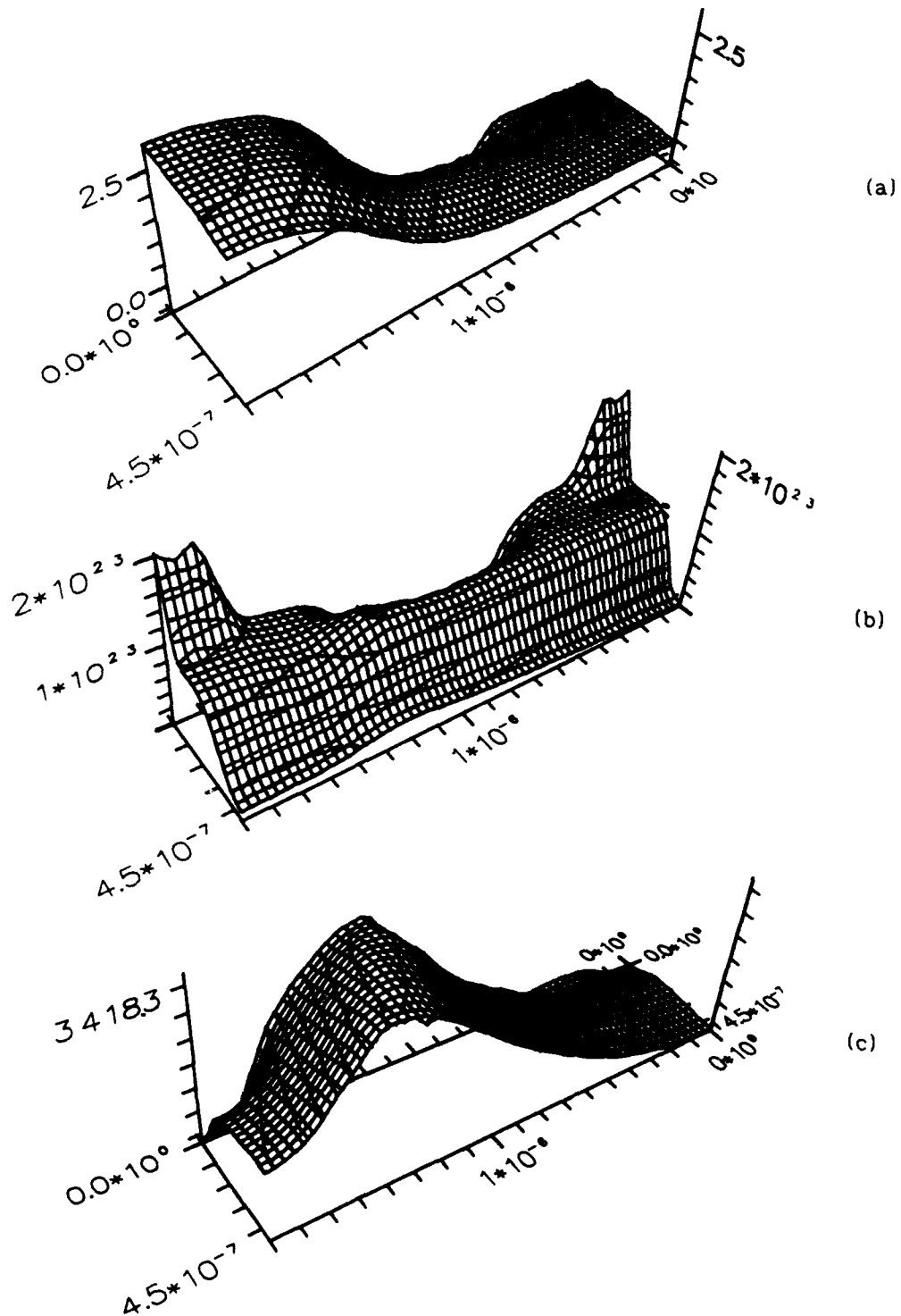


Fig.3. Steady-state results for (a) potential ψ , (b) electron density n , and (c) electron temperature T_e .

References

- [1] Selberherr S (1984) Analysis and simulation of semiconductor devices. Springer-Verlag, Vienna, New York
- [2] Mobbs SD (1989) Numerical techniques - finite element. Semiconductor Device Modelling (Snowden CM ed., Springer-Verlag, London, Berlin):49-59
- [3] Ingham DB (1989) Numerical techniques - finite difference and boundary element method. Semiconductor Device Modelling (Snowden CM ed, Springer-Verlag, London, Berlin):34-48
- [4] Xue H, Howes MJ and Snowden CM (1991) The modified semiconductor equations and associated algorithms for physical simulation. Int J Num Modelling: Electronic Networks, Devices and Fields 4:107-122
- [5] De Mari A (1968) An accurate numerical steady-state one-dimensional solution of the p-n junction. Solid State Electronics 11:33-58
- [6] Barton T (1989) Computer simulations. Semiconductor Device Modelling (Snowden CM ed, Springer-Verlag, London, Berlin):227-247
- [7] Cole EAB (1993) Numerical methods and their application to device modelling. Compound Semiconductor Device Modelling (Snowden CM and Miles RE eds, Springer-Verlag, London, Berlin):1-25
- [8] Scraton RE (1987) Further numerical methods. Edward Arnold, London, Baltimore
- [9] Varga RS (1962) Matrix iterative algebra. Prentice Hall Inc, New Jersey
- [10] Press WH, Flannery BP, Teukolsky SA and Vetterling WT (1986) Numerical Recipes. Cambridge University Press, Cambridge, London, New York
- [11] Kurata M (1982) Numerical analysis for semiconductor devices. Lexington Books, Lexington, Toronto
- [12] Smith GD (1985) Numerical solution of partial differential equations: finite difference methods. Clarendon Press, Oxford
- [13] Wilkinson JH and Reinsch C (1971) Handbook for automatic computation, vol 2. Springer-Verlag, Berlin
- [14] Young DM (1971) Iterative solution of large linear systems. Academic Press, New York
- [15] Stoer J and Bulirsch R (1980) Introduction to numerical analysis. Springer-Verlag, New York
- [16] Reiser M (1972) Large-scale numerical simulation in semiconductor device modelling. Computer Methods in App Mech and Eng 1:17-38
- [17] Stern F (1970) Iteration methods for calculating self-consistent fields in semiconductor inversion layers. J Comp Phys 6:56-67
- [18] Curtis A and Reid JK (1974) The choice of step-length when using differences to approximate Jacobian matrices. J Inst Math Appl 13:121-126
- [19] Broyden CG (1972) in Numerical methods for unconstrained optimisation (Murray W ed., Academic Press, New York)
- [20] Reiser M (1972) A two-dimensional numerical FET model for dc- ac- and large signal analysis. IBM Research J, April, RZ499
- [21] Snowden CM(ed) (1988) Semiconductor device modelling. Peter Peregrinus, London
- [22] Scharfetter DL and Gummel HK (1969) Large signal analysis of a silicon Read diode oscillator. IEEE Trans ED-16:64-77
- [23] McAndrew CC, Singhal K and Heasell EL (1985) A consistent nonisothermal extension of the Scharfetter-Gummel stable difference approximation. IEEE ED Lett EDL6:446-447
- [24] Blakemore JS (1982) Approximations for Fermi-Dirac integrals. Solid State Electronics 25:1067-1076
- [25] de Graaff HC and Klaassen FM (1990) Compact transistor modelling for circuit design. Springer-Verlag, Wien, New York
- [26] Stratton R (1972) Semiconductor current-flow. IEEE Trans ED-19:1288-1292

The Classical and Quantum Hydrodynamic Models

Carl L. Gardner¹

Department of Computer Science and Department of Mathematics
Duke University
Durham, NC 27706

Abstract

The classical and quantum hydrodynamic equations are presented in a unified formulation and the 3D transport equations are mathematically classified. The 1D steady-state classical and quantum equations are discretized in conservation form using an upwind method. A classical hydrodynamic simulation of a steady-state electron shock wave in a one micron Si semiconductor device at 77 K is presented and compared with a DAMOCLES simulation of the Boltzmann equation. Quantum hydrodynamic simulations of a resonant tunneling diode are presented which show charge buildup in the quantum well and negative differential resistance in the current-voltage curve.

I. Introduction

Electron propagation in a semiconductor crystal is well modeled down to submicron scales by the classical hydrodynamic model. The classical hydrodynamic equations can be extended to include quantum effects by incorporating the first quantum corrections. These $O(\hbar^2)$ terms allow particle tunneling through potential barriers and particle buildup in potential wells.

The aim of this paper is to give a unified presentation of the classical and quantum hydrodynamic conservation laws and of their mathematical classification and numerical discretization. I will also present a classical hydrodynamic (CHD) simulation of an electron shock wave in a one micron Si $n^+ - n - n^+$ device at 77 K and quantum hydrodynamic (QHD) simulations of resonant tunneling in an GaAs/Al_xGa_{1-x}As diode at 77 K.

The classical and quantum hydrodynamic conservation laws have the same form:

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial x_i}(nu_i) = 0 \quad (1)$$

$$\frac{\partial p_j}{\partial t} + \frac{\partial}{\partial x_i}(u_i p_j - P_{ij}) = -n \frac{\partial V}{\partial x_j} - \frac{p_j}{\tau_p} \quad (2)$$

$$\frac{\partial W}{\partial t} + \frac{\partial}{\partial x_i}(u_i W - u_j P_{ij} + q_i) = -nu_i \frac{\partial V}{\partial x_i} - \frac{(W - \frac{3}{2}nT_0)}{\tau_w} \quad (3)$$

¹Research supported in part by the U.S. Army Research Office under grant DAAL03-91-G-0146 and by the National Science Foundation under grant NSF-DMS-92-04189.

where n is the electron density, u is the velocity, p is the momentum density, P_{ij} is the stress tensor, $V = -e\phi$ is the potential energy, ϕ is the electric potential, $e > 0$ is the electronic charge, W is the energy density, q is the heat flux, and T_0 is the lattice temperature in energy units (k_B is set equal to 1). Indices i, j equal 1, 2, 3, and repeated indices are summed over. Eq. (1) expresses conservation of electron number, Eq. (2) expresses conservation of momentum, and Eq. (3) expresses conservation of energy. The collision terms in Eqs. (2) and (3) are modeled here by the standard relaxation time approximation, with momentum and energy relaxation times τ_p and τ_w .

The classical and quantum hydrodynamic equations can be derived from a moment expansion of the Wigner-Boltzmann equation. The classical equations are obtained by setting $\hbar = 0$. To close the moment expansion at three moments, we have to define e.g. p , P_{ij} , W , and q in terms of n , u , and T , where T is the electron temperature.

In the simplest approximation, the heat flux is specified by the Fourier law $q = -\kappa \nabla T$. For the $O(\hbar^2)$ "momentum-shifted" thermal equilibrium Wigner distribution function, the momentum density $p = mnu$, where m is the effective electron mass, the stress tensor is given by

$$P_{ij} = -nT\delta_{ij} + \frac{\hbar^2 n}{12m} \frac{\partial^2}{\partial x_i \partial x_j} \log(n) + O(\hbar^4) \quad (4)$$

and the energy density by

$$W = \frac{3}{2}nT + \frac{1}{2}mnu^2 - \frac{\hbar^2 n}{24m} \nabla^2 \log(n) + O(\hbar^4). \quad (5)$$

I derived the full three-dimensional quantum hydrodynamic model for the first time by a moment expansion of the Wigner-Boltzmann equation in Ref. [1]. The quantum correction to the energy density was first derived by Wigner [2]. The quantum correction to the stress tensor was proposed by Ancona and Tiersten [3] on general thermodynamical grounds and derived by Ancona and Iafrate [4] in the Wigner formalism. In the one-dimensional case, the 3D QHD equations reduce to the QHD model of Grubin and Kreskovsky [5].

The actual expansion parameter in the QHD equations is $\hbar^2/8mTl^2$, where l is a characteristic length scale of the problem [4]. For the resonant tunneling diode simulations in section V with $T \approx T_0 = 77$ K and $l = 100$ Å, the expansion parameter ≈ 0.23 .

The transport equations (1)–(3) are coupled to Poisson's equation for the electric potential energy

$$\nabla \cdot (\epsilon \nabla V) = e^2(N_D - N_A - n) \quad (6)$$

where ϵ is the dielectric constant, N_D is the density of donors, and N_A is the density of acceptors.

II. Classification of the Hydrodynamic Equations

To classify the hydrodynamic equations, rewrite the hydrodynamic equations (1)–(3) and (6) (with $\tau_p, \tau_w \rightarrow \infty$) in terms of n , u , and T :

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial x_i}(nu_i) = 0 \quad (7)$$

$$\frac{\partial u_j}{\partial t} + u_i \frac{\partial u_j}{\partial x_i} + \frac{1}{mn} \frac{\partial}{\partial x_j} (nT) - \frac{\hbar^2}{12m^2 n} \frac{\partial}{\partial x_i} \left(n \frac{\partial^2}{\partial x_i \partial x_j} \log(n) \right) + \frac{1}{m} \frac{\partial V}{\partial x_j} = 0 \quad (8)$$

$$\frac{\partial T}{\partial t} + u_i \frac{\partial T}{\partial x_i} + \frac{2}{3} T \frac{\partial u_i}{\partial x_i} - \frac{2}{3n} \frac{\partial}{\partial x_i} \left(\kappa \frac{\partial T}{\partial x_i} \right) + \frac{\hbar^2}{36mn} \frac{\partial}{\partial x_i} (n \nabla^2 u_i) = 0 \quad (9)$$

$$-\nabla(\epsilon \nabla V) + e^2 (N_D - N_A - n) = 0. \quad (10)$$

Then linearize the PDEs (7)–(10) with respect to a Fourier mode perturbation, and freeze coefficients. Set

$$\begin{bmatrix} n \\ u_i \\ T \\ V \end{bmatrix} = \begin{bmatrix} \bar{n} \\ \bar{u}_i \\ \bar{T} \\ \bar{V} \end{bmatrix} + e^{-\sigma t + i\mathbf{k} \cdot \mathbf{x}} \begin{bmatrix} \delta n \\ \delta u_i \\ \delta T \\ \delta V \end{bmatrix} \quad (11)$$

where $[\bar{n}, \bar{u}_i, \bar{T}, \bar{V}]$ is a solution of the hydrodynamic equations.

Next write the linearized hydrodynamic equations in terms of the symbol S of the linearized PDE system (7)–(10) as

$$-\text{diag}\{\sigma, \sigma, \sigma, \sigma, \sigma, 0\} [\delta n, \delta u_i, \delta T, \delta V] + S [\delta n, \delta u_i, \delta T, \delta V] = 0 \quad (12)$$

As $|\mathbf{k}| \rightarrow \infty$, $S - \text{diag}\{\sigma, \sigma \delta_{ij}, \sigma, 0\}$ has the form

$$\begin{matrix} & i \rightarrow \\ j \downarrow & \begin{bmatrix} ik \cdot u - \sigma & ik_i n & 0 & 0 \\ ik_j \frac{T}{mn} + ik_j k^2 \frac{\hbar^2}{12m^2 n} & (ik \cdot u - \sigma) \delta_{ij} & i \frac{k_j}{m} & i \frac{k_j}{m} \\ ik_i \frac{\hbar^2}{36mn} \nabla^2 u_i & \frac{2}{3} ik_i T - ik_i k^2 \frac{\hbar^2}{36m} & ik \cdot u + \frac{2}{3} k^2 \frac{\epsilon}{n} - \sigma & 0 \\ 0 & 0 & 0 & k^2 \epsilon \end{bmatrix} \end{matrix} \quad (13)$$

where i labels columns in the velocity perturbation δu_i and j labels rows in the velocity equation (8). I have dropped the bar over the solution $[n, u_i, T, V]$, and have separately kept the leading terms in \mathbf{k} for the limits $\hbar \rightarrow 0$ and $\kappa \rightarrow 0$.

The mathematical type of the PDE system is determined by the asymptotic eigenvalues σ of the symbol as $|\mathbf{k}| \rightarrow \infty$. We need only consider the upper 5×5 block S_s of the symbol, since the coupling of the transport equations (7)–(9) to Poisson's equation (10) only introduces the elliptic Poisson mode, and does not affect the modes of the transport equations.

There are three physically interesting cases to consider:

(1) $\hbar = 0$, $\kappa = 0$ (electrostatics). The eigenvalues of the symbol S_s and corresponding modes are

$$\begin{cases} i(\mathbf{k} \cdot \mathbf{u} \pm kc) & \text{hyperbolic} \\ i\mathbf{k} \cdot \mathbf{u} & \text{hyperbolic, multiplicity 3} \end{cases} \quad (14)$$

where $c = \sqrt{5T/3m}$. There are five nonlinear waves in classical electrostatics corresponding to the five hyperbolic modes: two shock waves and three contact waves. Two

contact waves can be labeled by a jump in the tangential velocity u_t across the wave, and one contact wave by a jump in the temperature T .

(2) $\hbar = 0$, $\kappa > 0$ (hydrodynamic model or electrogasdynamics with heat conduction). The eigenvalues of the symbol S_s and corresponding modes are

$$\begin{cases} i(\mathbf{k} \cdot \mathbf{u} \pm kc) & \text{hyperbolic} \\ i\mathbf{k} \cdot \mathbf{u} & \text{hyperbolic, multiplicity 2} \\ \frac{2}{3}k^2\kappa/n + i\mathbf{k} \cdot \mathbf{u} & \text{parabolic} \end{cases} \quad (15)$$

where $c = \sqrt{T/m}$. With heat conduction, there are four nonlinear waves [6] in the CHD model corresponding to the four hyperbolic modes: two shock waves and two contact waves. The two contact waves can be labeled by a jump in the tangential velocity u_t across the wave. The contact wave corresponding to a discontinuity in T has disappeared due to the parabolic heat conduction term $\nabla \cdot (\kappa \nabla T)$ in Eq. (3).

(3) $\hbar \neq 0$, $\kappa > 0$ (quantum hydrodynamic model). The eigenvalues of the symbol S_s and corresponding modes are

$$\begin{cases} \pm ik^2 \frac{\hbar}{\sqrt{12m}} & \text{Schrödinger} \\ i\mathbf{k} \cdot \mathbf{u} & \text{hyperbolic, multiplicity 2} \\ \frac{2}{3}k^2\kappa/n & \text{parabolic} \end{cases} \quad (16)$$

There are two contact discontinuities (in u_t) in the QHD model corresponding to the two hyperbolic modes. Note that two of the hyperbolic modes (which allow shock discontinuities to form) in the classical hydrodynamic model have become Schrödinger modes when the quantum corrections are included.

Well-posed boundary conditions for the 2D (and by extension 3D) classical hydrodynamic model are formulated in Ref. [7], assuming subsonic flow at the inflow and outflow boundaries. Here I will simply note that in one dimension, the CHD model (with heat conduction) has two hyperbolic modes, one parabolic mode, and one elliptic mode, and the QHD model (with heat conduction) has two Schrödinger modes, one parabolic mode, and one elliptic mode. Thus six boundary conditions are necessary for the CHD model and eight boundary conditions for the QHD model. Well-posed boundary conditions for the 1D CHD equations are $n = N_D$, and $T = T_0$ (or $\partial T / \partial x = 0$) at x_{\min} and x_{\max} , with a bias ΔV across the device: $V(x_{\min}) = T \log(n/n_i)$ and $V(x_{\max}) = T \log(n/n_i) + e\Delta V$, where n_i is the intrinsic electron concentration. For the 1D QHD equations, I add $\partial n / \partial x = 0$ at x_{\min} and x_{\max} .

III. The Second Upwind Method

I will discretize the 1D steady-state hydrodynamic equations using an upwind method adapted from computational fluid dynamics. Since the upwind method requires velocity values $u_{-\frac{1}{2}}, u_{\frac{1}{2}}, u_{\frac{3}{2}}, \dots, u_{N-\frac{1}{2}}, u_{N+\frac{1}{2}}$ at the midpoints of the elements I_i ($i = 1, \dots, N$)

connecting grid points $i-1$ and i , I will use a staggered grid for u . (I impose a computational Neumann boundary condition $du/dx = 0$ at $i = 0$ and $i = N$ to determine $u_{-\frac{1}{2}}$ and $u_{N+\frac{1}{2}}$.) The variables n , T , and V are defined at the grid points $i = 0, 1, \dots, N-1, N$. The boundary conditions specify n , T , and V (and dn/dx for the QHD model) at $i = 0$ and $i = N$.

In one dimension, the steady-state CHD and QHD models consist of the three nonlinear conservation laws for electron number, momentum, and energy, plus Poisson's equation:

$$\begin{bmatrix} f_n \\ f_u \\ f_T \\ f_V \end{bmatrix} = \frac{d}{dx} \begin{bmatrix} ug_n \\ ug_u \\ ug_T \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ h_u \\ h_T \\ h_V \end{bmatrix} + \begin{bmatrix} 0 \\ s_u \\ s_T \\ s_V \end{bmatrix} = 0 \quad (17)$$

where (set $\hbar = 0$ for the CHD model)

$$g_n = n \quad (18)$$

$$g_u = mnu \quad (19)$$

$$g_T = \frac{5}{2}nT + \frac{1}{2}mnu^2 - \frac{\hbar^2 n}{8m} \frac{d^2}{dx^2} \log(n) + nV \quad (20)$$

$$h_u = \frac{d}{dx}(nT) - \frac{d}{dx} \left(\frac{\hbar^2 n}{12m} \frac{d^2}{dx^2} \log(n) \right) + n \frac{dV}{dx} \quad (21)$$

$$h_T = -\frac{d}{dx} \left(\kappa \frac{dT}{dx} \right) \quad (22)$$

$$h_V = \epsilon \frac{d^2 V}{dx^2} \quad (23)$$

$$s_u = \frac{mnu}{\tau_p} \quad (24)$$

$$s_T = \left(\frac{3}{2}nT + \frac{1}{2}mnu^2 - \frac{\hbar^2 n}{24m} \frac{d^2}{dx^2} \log(n) - \frac{3}{2}nT_0 \right) / \tau_w \quad (25)$$

$$s_V = e^2(N_D - N_A - n) \quad (26)$$

Equations $f_n = 0$, $f_T = 0$, and $f_V = 0$ are enforced at the interior grid points $i = 1, \dots, N-1$, while equation $f_u = 0$ is enforced at the midpoints of the elements l_i , $i = 1, \dots, N$.

In the second upwind method, the advection terms $d(ug)/dx$ in Eq. (17) are discretized using second upwind differences²

$$\frac{d}{dx}(ug)_i \approx (u_{i+\frac{1}{2}}g_R - u_{i-\frac{1}{2}}g_L)/\Delta x \quad (27)$$

²The second upwind method is a conservative extension of the original first-order upwind method.

where

$$g_R = \begin{cases} g_i & (u_{i+\frac{1}{2}} > 0) \\ g_{i+1} & (u_{i+\frac{1}{2}} < 0) \end{cases}, \quad g_L = \begin{cases} g_{i-1} & (u_{i-\frac{1}{2}} > 0) \\ g_i & (u_{i-\frac{1}{2}} < 0) \end{cases} \quad (28)$$

and second-order central differences are used for h_u , h_T , h_V , and s_T .

I use Newton's method to linearize the discretized version of Eq. (17):

$$J \begin{bmatrix} \delta n \\ \delta u \\ \delta T \\ \delta V \end{bmatrix} = - \begin{bmatrix} f_n \\ f_u \\ f_T \\ f_V \end{bmatrix} = -f, \quad \begin{bmatrix} n \\ u \\ T \\ V \end{bmatrix} \leftarrow \begin{bmatrix} n \\ u \\ T \\ V \end{bmatrix} + t \begin{bmatrix} \delta n \\ \delta u \\ \delta T \\ \delta V \end{bmatrix} \quad (29)$$

where J is the Jacobian and t is a damping factor between 0 and 1, chosen to insure that the norm of the residual f decreases monotonically.

IV. CHD Simulation of the Electron Shock Wave

The nonlinear hyperbolic modes of the CHD model allow shock waves to develop. A steady-state electron shock wave in a semiconductor device was first simulated in Ref. [8]. The shock simulation has been confirmed by a Monte Carlo simulation of Laux of the Boltzmann equation using the DAMOCLES [9] program. The semiconductor device is an $n^+ - n - n^+$ Si diode at 77 K with 0.1 micron source and drain, with n^+ doping density $N_D = 10^{18} \text{ cm}^{-3}$, and a 1.0 micron channel, with n doping density $N_D = 10^{15} \text{ cm}^{-3}$.

For the momentum and energy relaxation times in the hydrodynamic model, I use modified Baccarani-Wordemann models:

$$\tau_p = \tau_{p0} \frac{T_0}{T} \quad (30)$$

$$\tau_w = \frac{\tau_p}{2} \left(1 + \frac{\frac{3}{2}T}{\frac{1}{2}mv_s^2} \right) \quad (31)$$

where the low-energy momentum relaxation time τ_{p0} is set equal to 1.67 picoseconds from the DAMOCLES data for 0.00995 eV electrons in homogeneous Si and $v_s = v_s(T_0)$ is the saturation velocity. For Si at 77 K, $m = 0.24 m_e$ and $v_s = 1.2 \times 10^7 \text{ cm/s}$.

The hydrodynamic and Boltzmann simulations agree remarkably well [10] when the amount of heat conduction in the hydrodynamic model is adjusted. The best fit for the thermal conductivity κ in the Wiedemann-Franz law for heat conduction

$$\mathbf{q} = -\kappa \nabla T, \quad \kappa = \kappa_0 \tau_{p0} n T_0 / m, \quad (32)$$

is given by $\kappa_0 = 0.05$.

Figs. 1 and 2 compare the hydrodynamic and DAMOCLES simulations of the 77 K electron shock wave in Si at a bias $\Delta V = 1$ volt. DAMOCLES calculates a current of 4500 amps/cm², very close to the hydrodynamic value of 4460 amps/cm². The velocity plot Fig. 1 most clearly shows the shock profile, which is spread out slightly due to the parabolic heat conduction term in the CHD model. The flow is supersonic at the velocity

peak just inside the channel, and subsonic at the end of the "velocity overshoot" wave where the velocity makes a "bend" to a constant value in the channel.

The DAMOCLES velocity exhibits a Mach 2.1 shock profile based on both internal evidence³ and comparison with the hydrodynamic simulation.

The type of velocity overshoot illustrated in Fig. 1 is always associated with a shock wave for the $n^+ - n - n^+$ diode. As the electrons enter the channel, the electron velocity increases rapidly to a peak value greater than the saturation velocity v_s . At the same time, the electron temperature falls slightly as the electrons overcome the small potential barrier at the source/channel junction. Thus the electron Mach number M near the velocity peak is greater than $v_s/c > v_s/c_0$, where $c = \sqrt{T/m}$ is the soundspeed at temperature T and c_0 is the soundspeed at the ambient temperature T_0 . For Si at 77 K, $v_s = 1.2 \times 10^7$ cm/s, $c_0 = 7.0 \times 10^6$ cm/s, and $M > 1.7$. On the other hand, the electron flow near the channel/drain is subsonic since $v \approx v_s$, while $T \gg T_0$, making $c \gg v_s$.

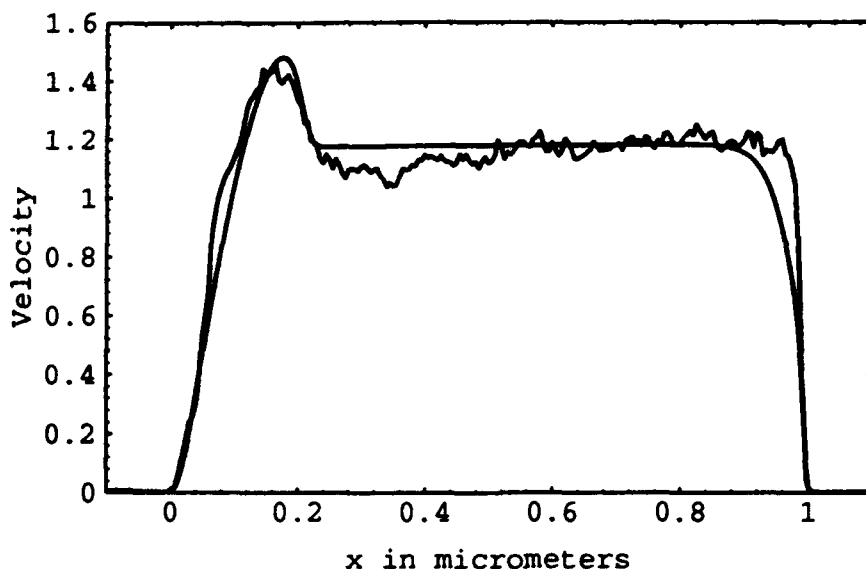


Figure 1: Hydrodynamic and Monte Carlo electron velocity in 10^7 cm/s for $\Delta V = 1$ volt. The jagged curve is the DAMOCLES result. The channel is between $x = 0$ and $x = 1$ micron.

The transition from supersonic flow to subsonic flow in general necessitates a shock wave in gas dynamics⁴—that is, a wave over which density, velocity, and (if heat conduction equals zero) temperature change very rapidly. The n^+ drain in the source-channel-

³The electron temperature $T \approx 77$ K at the shock wave. Using the effective electron mass approximation, the electron Mach number $M = v/c = v/\sqrt{T/m} \approx 2.1$.

⁴See Courant and Friedrichs [11], pp. 380-387.

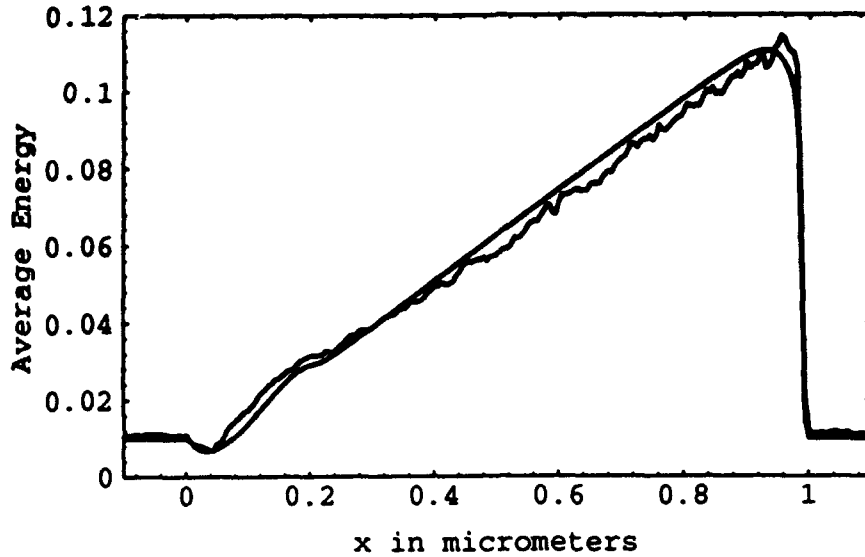


Figure 2: Hydrodynamic and Monte Carlo electron average energy in eV for $\Delta V = 1$ volt. The jagged curve is the DAMOCLES result.

drain structure of the diode provides the mechanism that forces a supersonic flow in the channel back down to subsonic flow.

The excellent agreement between the hydrodynamic and DAMOCLES results is remarkable in that the hydrodynamic model is orders of magnitude faster than Monte Carlo simulation of the Boltzmann equation. The hydrodynamic model also provides a mathematical framework in which to understand the velocity overshoot wave in the $n^+ - n - n^+$ diode.

V. QHD Simulation of the Resonant Tunneling Diode

The behavior of quantum devices that depend on particle tunneling through potential barriers and/or charge buildup in potential wells can be efficiently simulated using the QHD model. Here I will present simulations of a GaAs resonant tunneling diode with double $\text{Al}_x\text{Ga}_{1-x}\text{As}$ barriers, with barrier height $B = 0.209$ eV. The diode consists of an n^+ source (at the left) and an n^+ drain (at the right) with the doping density $N_D = 10^{18} \text{ cm}^{-3}$, and an n channel with $N_D = 5 \times 10^{15} \text{ cm}^{-3}$ (see Fig. 3). The channel is 250 Å long, the barriers are 50 Å wide, and the quantum well between the barriers is 50 Å wide. To enhance resonant tunneling, the device has 50 Å spacers between the barriers and the contacts.

The barrier height B is incorporated into the QHD transport equations (1)–(3) by replacing $V \rightarrow V + B$. (Poisson's equation is not changed.)

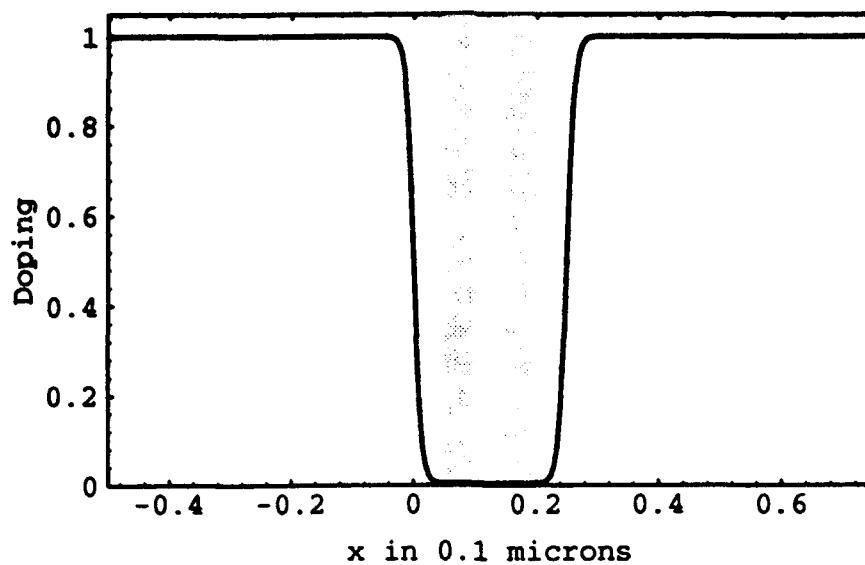


Figure 3: Doping/ 10^{18} cm^{-3} .

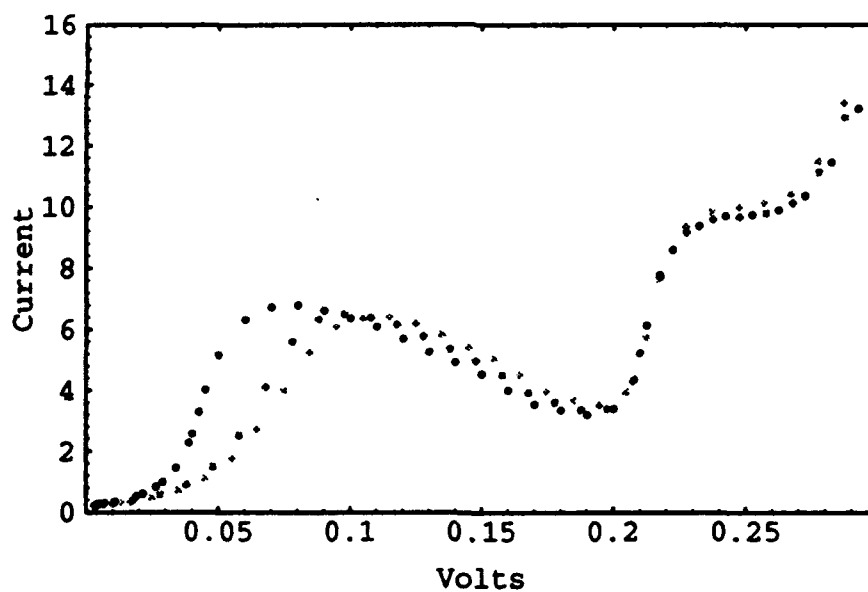


Figure 4: Current density in kiloamps/ cm^2 vs. voltage for the resonant tunneling diode at 77 K. $\kappa_0 = 0.2$ (black), 0.4 (dark gray), and 0.6 (gray). The dots represent computed solution points.

For lower valley electrons in GaAs at 77 K, the low-energy momentum relaxation time τ_{po} in Eqs. (30) and (31) is set equal to 0.9 picoseconds, the effective electron mass $m = 0.063 m_0$, and the saturation velocity (in Eq. (31)) $v_s \approx 2 \times 10^7$ cm/s. The dielectric constant $\epsilon = 12.9$ for GaAs.

Current-voltage curves for the resonant tunneling diode at 77 K are plotted in Fig. 4 for three different values of κ_0 in the expression for thermal conductivity (32). These are the first simulations [1] of the full QHD equations to show NDR in the resonant tunneling diode.

The peak of the current-voltage curve occurs as the electrons tunneling through the first barrier come into resonance with the the ground state of the quantum well. Note the presence of a "shoulder" in the current-voltage curve around $\Delta V = 0.25$ volts. The shoulder signals the location of the first virtual state of the quantum well. The location of the valley and shoulder can be qualitatively understood from the energy levels of a square well. For a 50 Å wide 0.209 eV high GaAs finite square well, there is just one bound state energy level at 0.079 eV. The energy of the first virtual state of the well is 0.24 eV.

The main effect of larger values of κ_0 is to shift the peak of the current-voltage curve to the right. With lower values of κ_0 , the electrons have a higher average energy as they impinge upon the first barrier, and therefore resonate with the well at a lower applied voltage.

A physically relevant value of κ_0 is approximately 0.4 for this device. The peak to valley current ratio of 1.95 agrees quantitatively with experimental ratios for similar devices.

Fig. 5 shows the dramatic charge enhancement in the quantum well typical of the resonant tunneling diode for applied voltages of $\Delta V = 0.097$ (peak), 0.191 (valley), and 0.22 volts (just before the shoulder) (with $\kappa_0 = 0.4$). The electron density at the center of the quantum well increases as ΔV increases, and is more than two orders of magnitude larger than the background doping density. Note the depletion of electrons around the channel-drain junction.

As illustrated in Fig. 6, the electrons spend the longest time in the quantum well for voltages near ΔV_{valley} . The "dwell" time spent by electrons in the well increases monotonically up to voltages near ΔV_{valley} , and then decreases rapidly. The macroscopic QHD dwell time differs qualitatively from dwell times based on microscopic quantum calculations, which predict that the dwell time is maximum at resonance.

Since the QHD equations have the same form as the classical fluid dynamical equations, well-understood classical boundary conditions can be applied in simulating quantum devices. Moreover, the QHD equations are expressed in terms of the fluid dynamical quantities density, velocity, and temperature. These classical fluid dynamical concepts enable us to interpret electron behavior in quantum devices in a physically intuitive way. We can define the time spent by an electron in the quantum well or the electron temperature throughout the device in a precise manner. In turn, the intuitive understanding developed through the QHD model sheds light on more fully quantum mechanical descriptions of electron behavior in quantum devices.

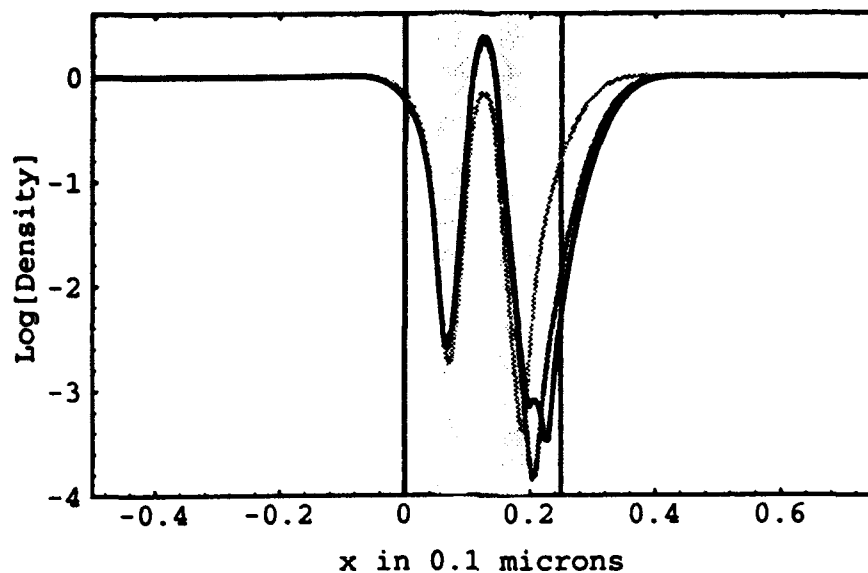


Figure 5: $\text{Log}[\text{Density}/10^{18} \text{ cm}^{-3}]$. The curves are for $\Delta V = 0.097$ (gray), 0.191 (dark gray), and 0.22 (black) volts.

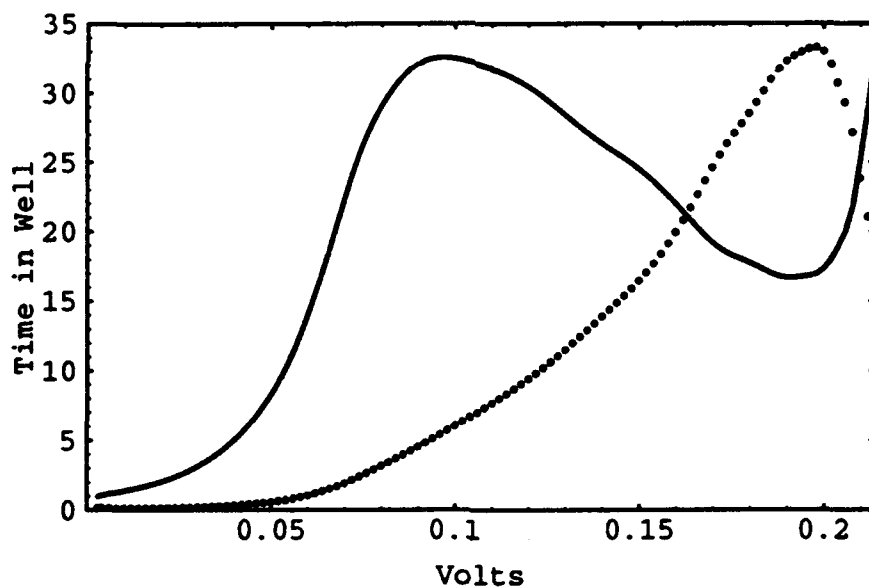


Figure 6: Time in picoseconds (dots) spent by electrons in the quantum well vs. voltage. For reference, a scaled version of the current density (solid line) is also shown.

References

- [1] C. L. Gardner, "The quantum hydrodynamic model for semiconductor devices," *SIAM Journal on Applied Mathematics*, to appear, 1993.
- [2] E. Wigner, "On the quantum correction for thermodynamic equilibrium," *Physical Review*, vol. 40, pp. 749-759, 1932.
- [3] M. G. Ancona and H. F. Tiersten, "Macroscopic physics of the silicon inversion layer," *Physical Review*, vol. B 35, pp. 7959-7965, 1987.
- [4] M. G. Ancona and G. J. Iafrate, "Quantum correction to the equation of state of an electron gas in a semiconductor," *Physical Review*, vol. B 39, pp. 9536-9540, 1989.
- [5] H. L. Grubin and J. P. Kreskovsky, "Quantum moment balance equations and resonant tunnelling structures," *Solid-State Electronics*, vol. 32, pp. 1071-1075, 1989.
- [6] C. L. Gardner, J. W. Jerome, and D. J. Rose, "Numerical methods for the hydrodynamic device model: Subsonic flow," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 8, pp. 501-507, 1989.
- [7] F. Odeh and E. Thomann, "On the well-posedness of the two-dimensional hydrodynamic model for semiconductor devices," *COMPEL*, vol. 9, pp. 45-47, 1990.
- [8] C. L. Gardner, "Numerical simulation of a steady-state electron shock wave in a sub-micrometer semiconductor device," *IEEE Transactions on Electron Devices*, vol. 38, pp. 392-398, 1991.
- [9] M. V. Fischetti and S. E. Laux, "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects," *Physical Review*, vol. B 38, pp. 9721-9745, 1988.
- [10] C. L. Gardner, "Hydrodynamic and Monte Carlo simulation of an electron shock wave in a one micrometer $n^+ - n - n^+$ diode," *IEEE Transactions on Electron Devices*, vol. 40, pp. 455-457, 1993.
- [11] R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves*. New York: Springer-Verlag, 1948.

Technology Computer-Aided Design

Siegfried Selberherr

Institute for Microelectronics
Technical University of Vienna
Gusshausstrasse 27-29, A-1040 Vienna, Austria

Abstract

Computer simulation is an indispensable tool for the design of new VLSI devices. Besides the progress in physical models and computational techniques which has dominated *technology computer aided design* (TCAD) in the past, purely software-related aspects are attracting increasing attention and will become a crucial issue for future developments in this field.

I. Introduction

Process and device simulation is commonly used for the design of new VLSI devices and processes and as an explorative tool to gain a better understanding of process and device physics. On the other hand, simulation is also carried out after the design phase in order to optimize certain parameters of a technology, e.g., to improve device performance and reliability or to increase the yield [1].

For all these tasks the term TCAD, short for *technology computer aided design*, was born. TCAD includes both physically rigorous and simplified process and device simulation in one to three spatial dimensions. Furthermore, links to layout-oriented CAD and to circuit simulation are required.

Depending on the particular application of TCAD tools, different demands arise: for the development of new technologies and for the prediction of the behavior of new devices both accuracy and robustness are required. In this case, very sophisticated physical models and numerical techniques must be used, usually at a high computational cost. An example of such a model is the two-dimensional simulation of the transient-enhanced diffusion during rapid thermal annealing [2].

On the other hand, for statistical simulations [3] or post-design process optimizations [4], speed is the most crucial issue, as physical models can be calibrated to an existing manufacturing process and hence do not pose a reliability problem.

Independently of the progress in advanced physical modeling, the fast and simple "tuned" models will still remain in broad use; there is no unique "best model" for all simulation problems.

TCAD involves a number of scientific disciplines in addition to electrical engineering and computer science. This has also had an impact on the properties of the software which has been produced by that heterogenous community during the past 20 years.

II. Frameworks

For a long time the importance of pure software issues for TCAD has been underestimated. In the past few years, as these issues attract increasing attention, the major focus is on the integration of TCAD tools into a common framework.

II.1 TCAD versus ECAD

In the *electronic CAD* (ECAD) field, frameworks for tool integration and data interchange standards have been emerging during the past years. Hardly any satisfying framework implementations are available for TCAD, especially for advanced, multi-dimensional process and device simulation. One reason for that is the difference in the sizes of the user communities. Electronic CAD is more widely used than Technology CAD.

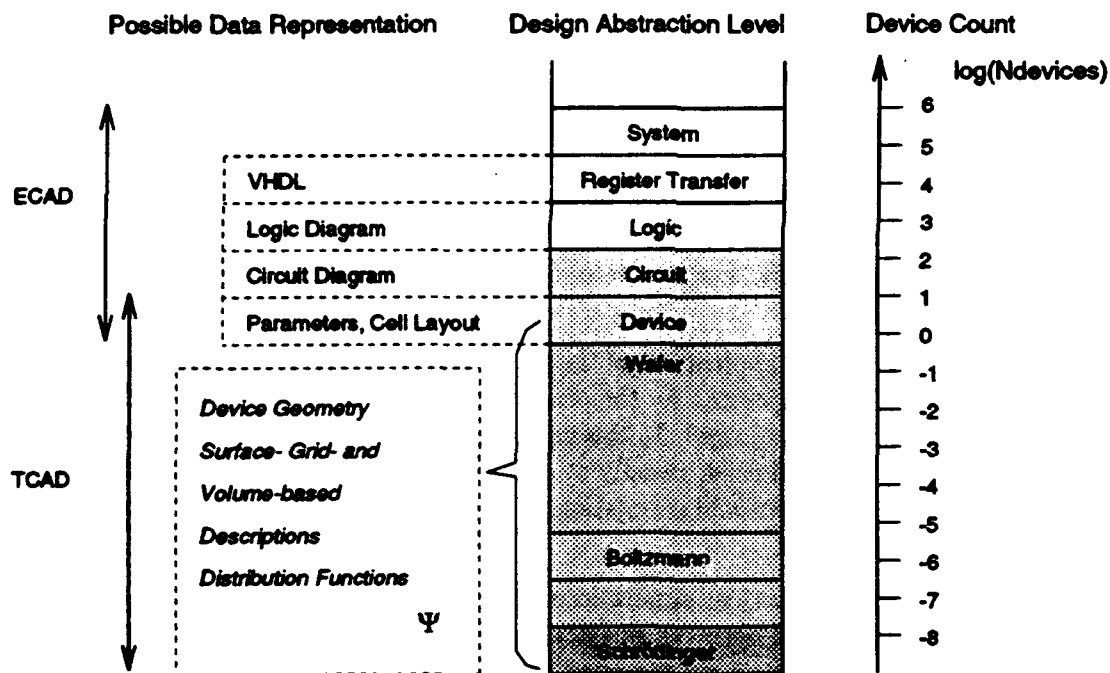


Figure 1: Design levels and device count scale in ECAD and TCAD

Another possible reason for the differences between ECAD and TCAD in terms of progress in frameworks and standardization is shown in Figure 1: in ECAD there have always been

several clearly defined layers of design abstraction. Though the device count scale is open towards the high end, there is a well-defined lower bound for ECAD, which is the single device. For every order of magnitude in device count there is a common abstraction level and at least one well-established data representation as well. For TCAD however, the only evident lower boundary in terms of abstraction is the physical atom. There are no "natural" intermediate abstraction levels and as an indirect consequence of physical modeling there is no well-established unique data representation like in the ECAD case.

These considerations indicate that finding a unified data representation for TCAD is dominated by semantic problems which are closely related to the large interval (Figure 1) of device count to be represented, together with the lack of clearly defined intermediate abstraction levels and the multi-disciplinary background involved in TCAD.

II.2 Framework Demands

As an operating environment for TCAD tools and engineers, a TCAD framework must provide the following key features:

- allow minimum effort integration of existing tools and facilitate the development of new tools
- allow casual users to use simulation in a black box manner
- provide enough flexibility on the task level to accomodate easily to new design tasks
- provide an extensible database for design representation
- be "open" in terms of platform independence, availability and the use of open standards
- provide standard functionality like visualization, interactive structure editing and postprocessing as generic tools

II.3 Existing Approaches

Facing these rather rigorous demands and the potential problems stated above, one cannot expect to find an easy and fast path leading to the ultimate TCAD framework. However, various attempts to head in that direction may be found:

In the semiconductor industry there have been emerging a number of remarkable framework efforts worldwide, such as an integrated system for statistical VLSI design [5] (Hitachi, Japan), an integrated, graphical device design environment [6] (Phillips, UK), SATURN [7] (Siemens, Germany) or the MECCA system [8] (AT&T, USA).

Based on an initial proposal by Duvall [9], there have been various *Profile Interchange Format* (PIF)-based design environments both in industry (like the PRIDE [10] system) and at universities (e.g. PROSE [11] from UC Berkely).

Commercial TCAD vendors are integrating their tools and providing them with unified user interfaces, like STUDIO from Technology Modeling Associates, or MASTERPIECE from Silvaco Data Systems.

Depending on the intentions of the creator of the frameworks, different aspects, like a rigorous task level implementation or a comfortable user interface [12], have been emphasized. Unfortunately, this has often been done at the expense of portability, or by leaving out a unified data representation.

Recently, a client-server framework architecture has been introduced by the semiconductor wafer representation working group [13] of the *CAD Framework Initiative* (CFI), an international standardization committee for ECAD. The intriguing goal of this approach is to separate the physical modeling completely from tedious tasks such as grid generation, interpolation, or geometry handling by providing these functions as a black-box server which is accessed by the simulation clients via a procedural interface. This method is very well-suited for the simulation of topography formation, however, it can be detrimental to applications with high data throughput or applications which exhibit performance advantages thanks to a tight coupling between physical models and numerical techniques.

In an attempt to address all of the framework demands stated above, we have developed VISTA, the *Viennese Integrated System for TCAD Applications*. It consists of a PIF Database, which is an enhanced intertool version of the well-known profile interchange format proposed in [9]. To accomodate the needs of existing TCAD applications, the original PIF syntax was restructured by reducing the number of different constructs, adding a few new constructs such as tensor product grid definition, and by defining additional semantic rules for the use of standardized attributes. Our PIF implementation can be used to store arbitrary LISP expressions for process flow representation.

Simulators and all other tools access the PIF database using the *PIF Application Interface* (PAI) [14], which supports several programming languages including C, FORTRAN and LISP. The PAI is a procedural interface for accessing the binary PIF database. It provides functionality for creating, reading and modifying PIF objects. In this way the application programmer does not need to know too much about the PIF syntax to be able to use the PIF database. The PAI was designed as a strictly layered product to guarantee the necessary functionality, performance and extensibility.

A *system layer* hides all system dependencies concerning communication with the operating system from the rest of the PAI. A *caching layer* takes care of performance and space requirements. The *interface layer* allows access to the PIF objects suited for advanced C and is the standardized interface to the PIF database. The application layer provides a more comfortable access to PIF objects for applications written in C, FORTRAN or LISP.

For the conversion from the binary intertool form to the intersite ASCII format the PIF binary file manager (PBFM) has been developed. In this way the intersite exchange of PIF files between different hardware and software platforms, for instance via electronic mail, is supported.

The simulators are controlled by an interpreting TCAD shell [15] which integrates all system components on the task level. The major benefits of selecting LISP as extension

language are that full-fledged programming language features like branches, loops and subprograms, mechanisms for defining new variables and standard mathematical operations and expressions can be utilized to define complex development tasks. The XLISP [16] interpreter which we have chosen as basis for the TCAD shell, is a publicly available software product written in portable C, and has the additional feature of being highly system-independent.

The user interface of VISTA [17] is based on the X Window system. By means of the X Toolkit [18] it implements most of the functionality which is required to support TCAD information flow as so-called *widgets*. The user interface is tightly coupled with the XLISP interpreter. This combination of a widget-oriented user interface with an interpreter is known to be a very flexible and promising concept [19], [20], [21].

An interactive device editor, generic postprocessing and visualization modules are also part of the framework. Special emphasis has been put on the use of open portable subsystems (which are mostly public domain products) to achieve a high degree of portability.

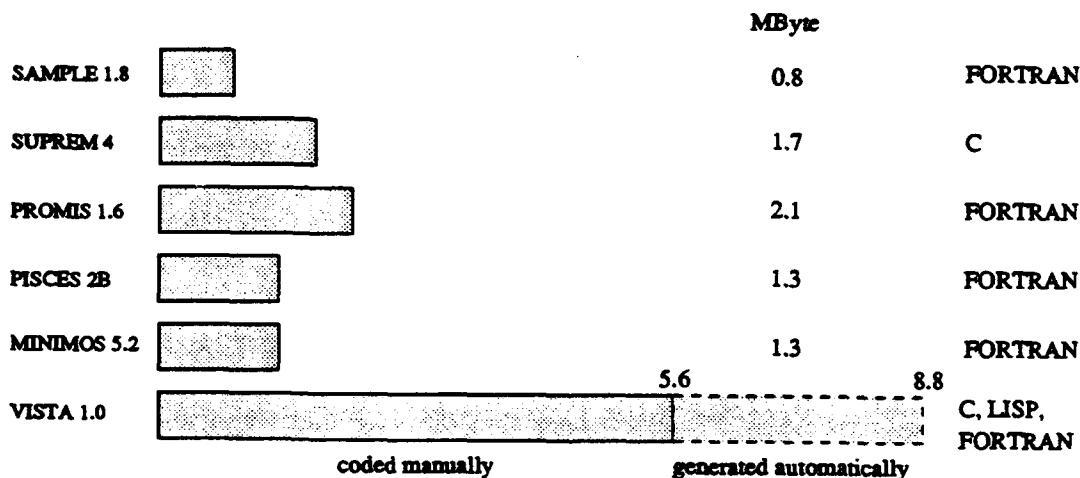


Figure 2: Comparison of source code sizes. VISTA does not include simulation tools.

Figure 2 shows the code sizes of well-known simulation tools. SAMPLE [12] (from UC Berkeley) is a topography simulation tool, MINIMOS [22] and PROMIS [2] are two-dimensional MOS device and process simulation programs, respectively, developed at TU Vienna. SUPREM4 is a two-dimensional process simulation program and PISCES is a two-dimensional device simulation program, both from Stanford University [23].

Given VISTA's remarkable code size compared to these classical single simulation tools, it becomes obvious that a framework for tool integration and development must exhibit a homogenous architecture in order to be comprehensible and maintainable.

This can be achieved through the use of unified concepts. The generalization of existing solutions within the framework should be favoured over the introduction of new ones. A second key to a comprehensible system is the use of abstract concepts. In VISTA, tedious

parts of the sourcecode, like the PAI or interface code which is required for multilanguage programming, is generated automatically from abstract specifications. Only the creative parts are left to the programmer.

III. The Future

It is a fact that the rapidly shrinking device dimensions and the increasing cost of experimental design optimization have been stimulating the evolution of TCAD; it can be expected that TCAD will continue to evolve into several directions:

The exact future development of physical models is hard to predict, as models always reflect the state of technology [24], [25]. Future models will have to be based on a lower level of abstraction. As a consequence the complexity and the "bandwidth" of models used for TCAD will increase. As to computational techniques, it is obvious that the availability of massively parallel computers will stimulate the development and use of parallelizable methods, like Monte Carlo simulation, e.g. [26].

Besides the progress in physical modeling, the future of process and device simulation will be significantly influenced by the introduction of TCAD frameworks and its impacts. This integration will allow simulation to catch up with the physical reality. It is to be hoped that TCAD frameworks will help to reduce the gap between the engineer's simulation needs and the sophistication of the available models and simulation methods by providing a "plug-and-play" environment for tool developers. One major requirement towards this goal is the semantic standardization for TCAD data, which concerns describing process information in a comprehensive and unambiguous way rather than finding appropriate representation methods. Client-server concepts can be expected to be used in future simulation tools in different places, e.g. for solving large linear systems.

A new object-oriented method for CAD tool management has been demonstrated through the Cadwell design framework [27]. We believe that a similar method of tool abstraction should be employed in a framework for advanced process and device simulation as well.

The importance of process and device simulation will increase in general, and as a consequence of the broader use, performance, together with robustness and ease of use will become even more crucial. The integration of different process and device simulators into modern TCAD systems will become a necessity because of the inherent communication, data transfer and maintainance advantages.

Acknowledgement

The VISTA project is supported by the "Forschungsförderungsfonds für die gewerbliche Wirtschaft", project 2/285 and 2/299, and by the research laboratories of AUSTRIAN INDUSTRIES - AMS at Unterpemstätten, Austria; DIGITAL EQUIPMENT at Hudson, USA; HITACHI at Tokyo, Japan; PHILIPS at Eindhoven, the Netherlands; SIEMENS at Munich, FRG; and SONY at Atsugi, Japan.

References

- [1] D. Cole, E. Buturla, S. Furkay, K. Varahramyan, J. Slinkman, J. Mandelman, D. Foty, O. Bula, A. Strong, J. Park, T. Linton Jr., J. Johnson, M. Fischetti, S. Laux, P. Cottrell, H. Lustig, F. Pileggi, and D. Katcoff, "The Use of Simulation in Semiconductor Technology Development," *Solid-State Electron.*, Vol. 33, No. 6, pp. 591-623, June 1990.
- [2] G. Hobler, S. Halama, W. Wimmer, S. Selberherr, and H. Pötzl, "RTA-Simulations with the 2-D Process Simulator PROMIS," in Proc. *NUPAD III Workshop*, (Honolulu), pp. 13-14, 1990.
- [3] J. Brockman and S. Director, "A Macromodeling Approach to Process Simulator Tuning," in Proc. *NUPAD III Workshop*, (Honolulu), pp. 17-18, 1990.
- [4] J. Wenstrand, H. Iwai, M. Norishima, H. Tanimoto, T. Wada, and R. Dutton, "Intelligent Simulation for Optimization of Fabrication Processes," in Proc. *NUPAD III Workshop*, (Honolulu), pp. 15-16, 1990.
- [5] H. Matsuo, H. Masuda, S. Yamamoto, and T. Toyabe, "A Supervised Process and Device Simulation for Statistical VLSI Design," in Proc. *NUPAD III Workshop*, pp. 59-60, 1990.
- [6] P. Gough, M. Johnson, P. Walker, and H. Hermans, "An Integrated Device Design Environment for Semiconductors," *IEEE Trans. Computer-Aided Design*, Vol. 10, No. 6, pp. 808-821, June 1991.
- [7] H. Jacobs, W. Hänsch, F. Hofmann, W. Jacobs, M. Paffrath, E. Rank, K. Steger, and U. Weinert, "SATURN - a Device Engineer's Tool for Optimizing MOSFET Performance and Lifetime," in Proc. *NUPAD III Workshop*, (Honolulu), pp. 55-56, 1990.
- [8] P. Lloyd, H. Dirks, E. Prendergast, and K. Singhal, "Technology CAD for Competitive Products," *IEEE Trans. Computer-Aided Design*, Vol. 9, No. 11, pp. 1209-1216, November 1990.
- [9] S. Duvall, "An Interchange Format for Process and Device Simulation," *IEEE Trans. Computer-Aided Design*, Vol. 7, No. 7, pp. 741-754, July 1988.
- [10] M. Simpson, "Pride: An Integrated Design Environment for Semiconductor Device Simulation," *IEEE Trans. Computer-Aided Design*, Vol. 10, No. 9, pp. 1163-1174, September 1991.
- [11] A. Wong, *Technology Computer-Aided Design Frameworks and the PROSE Implementation*. PhD thesis, University of California, Berkeley, 1992.
- [12] E. Scheckler, A. Wong, R. Wang, G. Chin, J. Camanga, A. Neureuther, and R. Dutton, "A Utility-Based Integrated System for Process Simulation," *IEEE Trans. Computer-Aided Design*, Vol. 11, No. 7, pp. 911-920, July 1992.

- [13] D. Boning, G. Chin, R. Cottle, W. Dietrich, S. Duvall, M. Giles, R. Harris, M. Karasick, N. Khalil, M. Law, M.J. McLennan, P. Mozumder, L. Nackman, S. Nassif, V. Rajan, D. Schröder, R. Tremain, D. Walker, R. Wang, and A. Wong, "Developing and Integrating TCAD Applications with the Semiconductor Wafer Representation," in *Proc. NUPAD IV Workshop*, (Seattle), pp. 199-204, 1992.
- [14] F. Fasching, C. Fischer, S. Selberherr, H. Stippel, W. Tuppa, and H. Read, "A PIF Implementation for TCAD Purposes," in *Proc. SISDEP 91 Conf.*, (Zürich), pp. 477-482, 1991.
- [15] H. Pimingstorfer, S. Halama, S. Selberherr, K. Wimmer, and P. Verhas, "A Technology CAD Shell," in *Proc. SISDEP 91 Conf.*, (Zürich), pp. 409-416, 1991.
- [16] D. Betz, *XLISP: An Object-Oriented LISP, Version 2.1*, 1989.
- [17] S. Halama, F. Fasching, H. Pimingstorfer, W. Tuppa, and S. Selberherr, "Consistent User Interface and Task Level Architecture of a TCAD System," in *Proc. NUPAD IV Workshop*, (Seattle), pp. 237-242, 1992.
- [18] P. Asente and R. Swick, *X Window System Toolkit, The Complete Programmer's Guide and Specification*. Digital Press, 1990.
- [19] N. Mayer, "WINTERP: An Object-Oriented Rapid Prototyping, Development and Delivery Environment for Building User-Customizable Applications with the OSF/Motif UI Toolkit.," Tech.Rep., Hewlett-Packard Laboratories, Palo Alto, 1991.
- [20] J. Ousterhout, "TCL: An Embeddable Command Language," in *Proc. 1990 Winter USENIX Conf.*, (Dallas), pp. 133-146, 1990.
- [21] J. Ousterhout, "An X11 Toolkit Based on the TCL Language," in *Proc. 1991 Winter USENIX Conf.*, (Dallas), pp. 105-115, 1991.
- [22] S. Selberherr, "Three-Dimensional Device Modeling with MINIMOS 5," in *Proc. VPAD 89 Workshop*, (Osaka), pp. 40-41, 1989.
- [23] J. Pfister, L. Parrillo, and F. Baker, "A Physical Model for Boron Penetration Through Thin Gate Oxides from p⁺ Polysilicon Gates," *IEEE Electron Device Lett.*, Vol. 11, No. 6, pp. 247-249, June 1990.
- [24] M. Lundstrom and S. Datta, "Physical Device Simulation in a Shrinking World," *IEEE Circuits and Devices Mag.*, Vol. 6, No. 4, pp. 32-37, July 1990.
- [25] M. Orlowski, "Challenges for Process Modeling and Simulation in the 90's - an Industrial Perspective," in *Proc. SISDEP 91 Conf.*, (Zürich), pp. 3-22, 1991.
- [26] S. Laux, M. Fischetti, and D. Frank, "Monte-Carlo Analysis of Semiconductor Devices: The DAMOCLES Program," *IBM J.Res.Dev.*, Vol. 34, No. 4, pp. 466-494, July 1990.
- [27] J. Daniell and S. Director, "An Object-Oriented Approach to CAD Tool Control," *IEEE Trans.Computer-Aided Design*, Vol. 10, No. 6, pp. 698-713, June 1991.

The Application of Quadrilateral Finite Elements for the Simulation of Recess T-gate MESFETs and HEMTs

Asen Asenov, Don Reid, John R. Barker, Nigel Cameron, and Steve P. Beaumont

Nanoelectronics Research Centre,
Department of Electronics and Electrical Engineering
University of Glasgow,
Glasgow, G12 8QQ, UK

Abstract

In this paper we present a new finite element approach for the simulation of recessed T-gate MESFETs and HEMTs based on quadrilateral finite elements. The discretization of the current continuity equation, which is the crucial part of the simulation, is followed in details. Two simulation examples - a 200 nm gate length MESFET and a δ -doped pseudomorphic HEMT illustrates the usefulness of the adopted approach.

I. Introduction

The performance of the modern nanometer-scale MESFETs and HEMTs becomes strongly affected by device parasitics such as coupling capacitances and access resistances [1]. In recessed gate devices these parasitics are critically affected by the shape and surface condition of the recess region. In addition the T-gate process designed to reduce the gate series resistance [2] may also reinforce the parasitic capacitances. Although Hydrodynamic [3] and Monte Carlo [4] simulation programs are making significant progress in properly describing the non-equilibrium transport phenomena in compound FETs, the real shape of the gate recess is generally poorly modelled, assuming planar or rectangular simulation domains. Surface effects are also either neglected or modelled by fixing the surface potential or by increasing the surface doping. Yet it is well known that this effects can in many cases have a more profound impact on device DC characteristics and high frequency performance than the transport details in the 'intrinsic' region under the gate.

In this paper we describe the implementation of a finite element approach based on quadrilateral finite elements for a precise description of the device's geometry and the realistic inclusion of the surface effects in the simulation of recessed gate MESFETs and HEMTs. Several simulation examples illustrate the work of the developed on this basis Heterojunction 2D Finite element (H2F) simulator.

II. Model Description and Discretization

The heterojunction compound semiconductor device model equations used in the current version of H2F for steady state simulation include Poisson's equation and electron current continuity equation in a drift diffusion approximation

$$\nabla \cdot [\epsilon(\bar{r}) \nabla \psi] = -\rho(\bar{r}) \quad (1)$$

$$\nabla \cdot \bar{J}_n = 0 \quad (2)$$

$$\bar{J}_n = -e\mu_n \nabla \psi_1 + eD_n \nabla n \quad (3)$$

$$\psi_i = \psi + \frac{\chi}{q} + \frac{kT}{q} \ln(N_c) \quad (4)$$

where all symbols have their usual meaning. While this approach is unable to describe precisely the device's transport, in many cases it is justified by the need to accurately predict the device parasitics.

A great deal of attention has been paid to the proper handling of the surface effects in the simulation. The charge term in the Poisson's equation is given by

$$\rho(\bar{r}) = q(N_{\text{bulk}}^i + \delta(\perp)N_{\text{int}}^i + p - n) \quad (5)$$

Where N_{bulk}^i and N_{int}^i represent charged bulk and interface states respectively. The symbol $\delta(\perp)$ indicates that N_{int}^i is surface charge which may be placed along any (in the general case curvilinear) interface in the device. Similar to the bulk model the generalised interface trap model includes acceptor and donor like states with an arbitrary energy position whose occupation depends on the quasi-Fermi level and the potential variations. The simulation domain includes the space above the semiconductor surface providing a proper interaction between the charge on the surface states and the spreading surface potential.

Quadrilateral finite elements have been used for the discretization. The grid is generated by appropriate deformation of originally rectangular sub domains. Fig. 1 a,b illustrates this procedure for the gap between the gate and the cap layer of a recessed gate FET

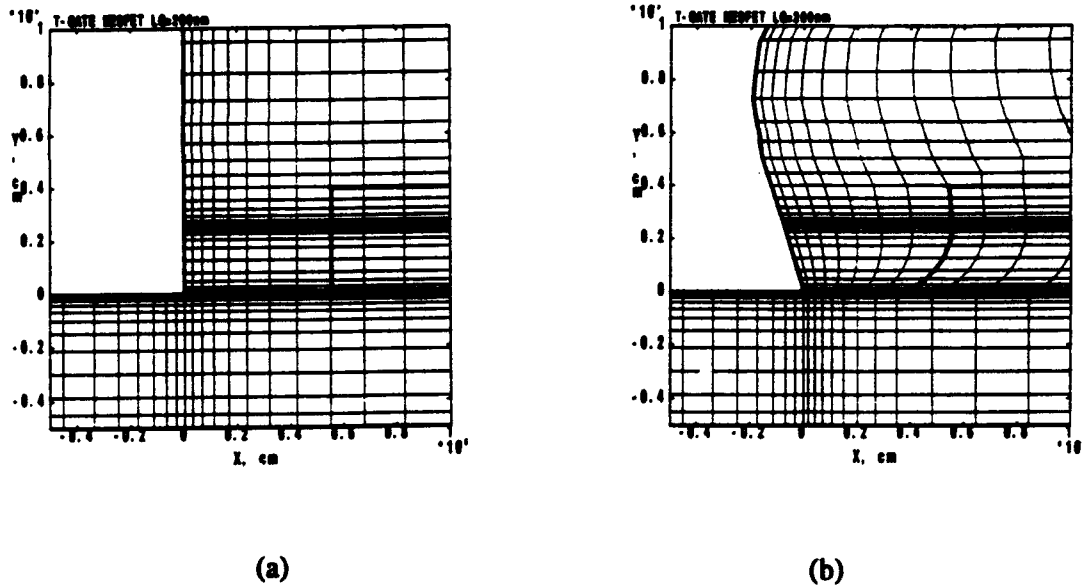


Fig. 1 Generation of the quadrilateral grid in the recess gate region. (a) the initial rectangular grid (b) the quadrilateral grid after suitable deformation

The Galerkin finite element method has been adapted to solve the Poisson's equation. The corresponding integration over the quadrilateral elements during the discretization was carried out by a linear isoparametric mapping (Fig. 3). A control volume method has been adopted [5] for the discretization of the current-continuity equation (Fig. 4). In this approach each quadrilateral element is divided into four subelements and the discretization is carried out balancing the current flowing in and out of the subelements attached to a given condensation point. Thus for vertex 1 (See Fig. 4) $I_{41} - I_{21} = 0$. A central point integration is used to calculate

the current through each of the subelement sides. For example $I_{12}=d_1J_{12}$, where the current density J_{12} is approximated by the standard Gummel expression.

$$J_{12} = \frac{D_{AB}}{l_1} \left[\text{ber} \left(\frac{q\Delta\psi_{AB}}{kT} \right) n_A - \text{ber} \left(\frac{-q\Delta\psi_{AB}}{kT} \right) n_B \right] \quad (6)$$

The growth functions involved in the derivation of this expression are also used for interpolation of the electron concentration along the sides of the element. The electron concentration in point A for example is given by

$$n_A = w_{A1}n_1 + w_{A4}n_4 \quad (7)$$

where

$$w_{A4} = \frac{x_{4A} \text{ber} \left(\frac{q\Delta\psi_{14}}{kT} \right)}{\text{ber} \left(\frac{q\Delta\psi_{14}x_{4A}}{kTl_4} \right)} \quad (8)$$

and

$$w_{A1} = 1 - w_{A4} \quad (9)$$

It has been found that this discretization is stable for arbitrary shapes of the quadrilateral elements and does not lead to the spikes typical for obtuse triangles. The same approach may be used for discretization of the momentum and energy conservation equation in the quasi-hydrodynamic transport treatment.

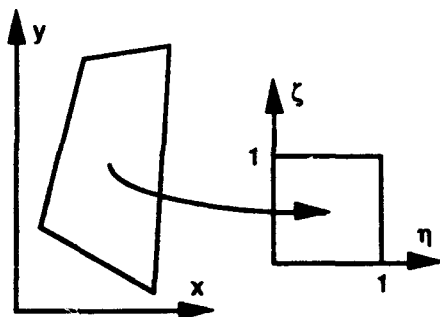


Fig. 2 Linear isoparametric mapping

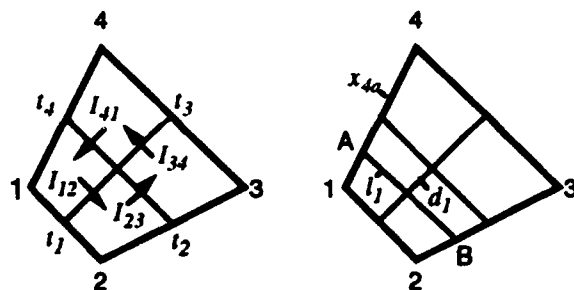


Fig. 3 Discretization of the current continuity equation

The grid generation preserves the number of grid points in lateral and vertical directions and leads to a regular nine diagonal matrix of the discretized equations. A Fast Incomplete LU Factorisation Biconjugate Gradients (ILUBCG) solver is used for the numerically intensive iterations. The solution of the Poisson's equation involves only a few biconjugate gradient steps per Newton iteration that significantly reduces the total computation time. The convergence problems related to the strongly localised, potential dependent interface charge have been resolved by appropriate dumping. ILUBCG also solves without complication the discretized current continuity equation.

Although H2F is 'serial' code, by using a pipeline fileserver, multiple copies of the program can be run concurrently on MIMD machines (in our case a Parsytec Model 64 transputer system), calculating in parallel, separate sets of input device parameters. This extends dramatically the capability of the simulator for real design work such as structure optimisation, sensitivity analysis and yield prediction where several hundred simulations are often carried out during single investigation.

III. Examples

Two simulation examples illustrate the application of the developed quadrilateral finite element approach for the simulation of complex shaped devices. The first example is concerned with the simulation of a 200 nm gate-length state of the art MESFET, fabricated in the Nanoelectronics Research Centre at the University of Glasgow [6]. The flexibility of the quadrilateral grid is illustrated in Fig. 4 where the cross sectional photograph of the device is compared with the corresponding H2F simulation domain.

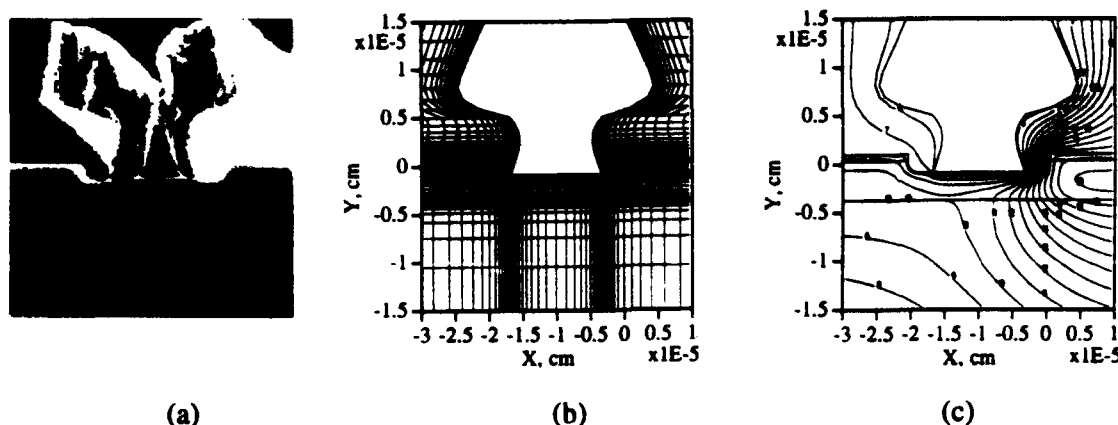


Fig. 4 Simulation of a 200 nm gate length MESFET. (a) Cross sectional SEM view (b) the corresponding finite element grid (c) potential distribution at $V_G = -0.4V$ and $V_D = 2.5V$

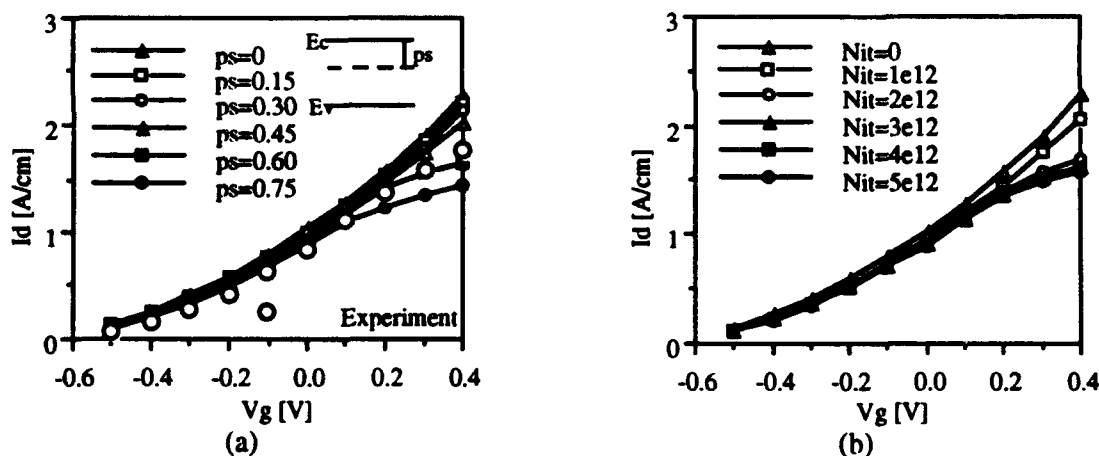


Fig. 5 Simulated and measured I_D - V_D characteristics of the device illustrated in Fig. 4. (a) influence of acceptor type surface states position P_s (b) influence of the surface state density N_{it}

The influence of the position and the density of these surface states on the device's I_D - V_G characteristics for the MESFET shown in Fig. 4 is given in Fig 5 a,b. Our experimental measurements are in good agreement with the expected position and states density $P_s = 0.6$ eV and $N_{it} = 2e12$ cm⁻².

The second example is the simulation of a delta-doped pseudomorphic HEMT whose behaviour is significantly influenced by the series resistances introduced by an undoped cap layer (Fig. 6). Although the drift diffusion approach underestimates the current, it has been found that by

adjusting the saturation velocity in the mobility model (to 1.4×10^7 cm/s in this case) the measured characteristics can be acceptably matched.

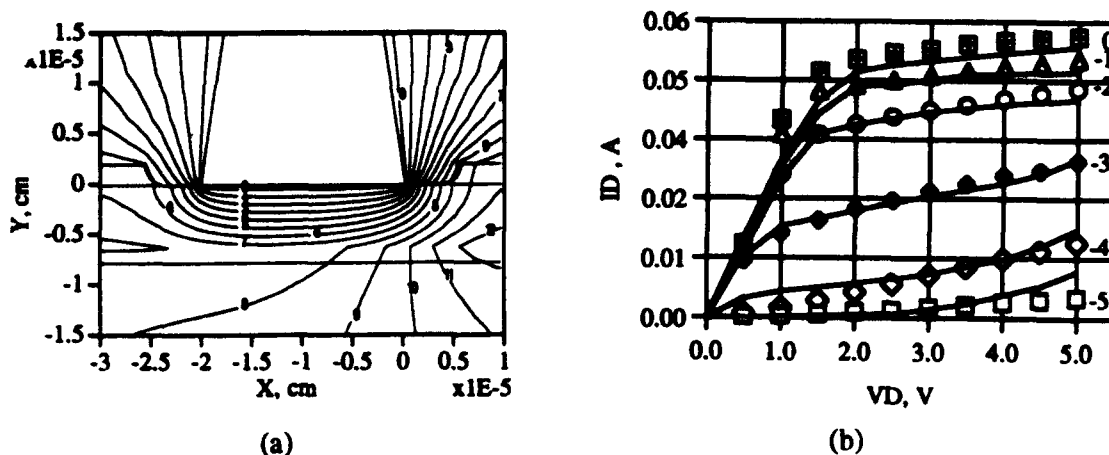


Fig. 6 Simulation of 200 nm gate length pseudomorphic HEMT (a) device structure and potential distribution at $V_G = 3.5$ V and $V_D = 2.5$ V (0 correspond to -4 V and the increment is 0.5 V) (b) measured (lines) and calculated (symbols) $I_D - V_G$ characteristics

IV. Conclusions

The developed quadrilateral finite element approach provides the necessary flexibility for the proper handling of shape and surface associated parasitic effects in the modern recessed T-gate FETs. In the example simulations, where device parasitics play an important role, the implemented drift-diffusion approach leads to a reasonable prediction of the device behaviour with a simple tuning of the saturated carrier velocity. However, the proposed finite element scheme may be extended to incorporate the hydrodynamic approach in the same manner as the finite difference case.

References

- [1] P.H. Ladbrooke, A.J. Hill and J.P. Bridge, "Fast FET and HEMT Solvers for Microwave CAD", J. Microwave and Millimetre-Wave Computer Aided Eng., Vol. 3, No. 1, pp.37-60, 1993.
- [2] P.C. Chao, P.M. Smit, S.C. Plamaeteer, and J.C.M. Hwang, "Electron Beam Fabrication of GaAs Low-Noise MESFET's Using a New TriLaur Resist Technique", IEEE Trans. Electron Dev., Vol. ED-32, No. 6, pp.1042-1045, 1985.
- [3] J.-R. Zhou and D.K. Ferry, "Simulation of Ultra-Small GaAs MESFET Using Quantum Moment Equations", IEEE Trans. Electron Dev., Vol 39, No. 3, pp.473-478, 1992.
- [4] I.C. Kizilyalli, M. Artaki and A. Chandra, "Monte Carlo Study of GaAs/ $\text{Al}_x\text{Ga}_{1-x}\text{As}$ MODFET's: The Effect of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ composition", IEEE Trans. Electron Dev., Vol. 38, No. 2, pp.197-206, 1991.
- [5] A. Asenov and E. Stefanov, "IMPEDANCE 2.0: A Flexible Concept for Process and Device Simulation", Proc. ISPPM, pp.272-286, Varna 1989.
- [6] N.I. Cameron, G. Hopkins, I.G. Thayne, S.P. Beaumont, C.D.W. Wilkinson, M. Holland, A.H. Keanand and C.R. Stanley, "Selective Reactive Ion Etching of GaAs/ AlGaAs Metal-Semiconductor Field Effect Transistors", J. Vac Sci. Technol. B, Vol. 9 p.3538, 1991.

Numerical Simulation of IGBTs at Elevated Temperatures

Andrew R. Brown[†], Asen Asenov[†], John R. Barker[†], Steve Jones^{*} and Peter Waind^{*}

[†]Department of Electronics and Electrical Engineering,
University of Glasgow, Glasgow G12 8QQ, UK

^{*}GEC Plessey Semiconductors,
Carholme Road, Lincoln LN1 1SG, UK

Abstract

In this paper we present a 2D finite difference device simulator which has been developed to underpin the design and optimisation of power IGBTs. The simulator has been used to investigate the effect of elevated temperatures on the performance of the IGBT for use in environments where ambient temperatures up to 200°C are likely. The preliminary results of this investigation are reported here.

I. Introduction

The Insulated Gate Bipolar Transistor (IGBT) (Fig.1) is at present one of the most widely used power semiconductor devices [1]. Its popularity among the designers of power electronics systems is due to the lucky combination of a low forward voltage drop, typical for power BJTs, and gate controlled turn-off and high switching speed characteristic of power MOSFETs [2]. At the same time the coupled field effect and bipolar actions make the device operation quite complicated [3].

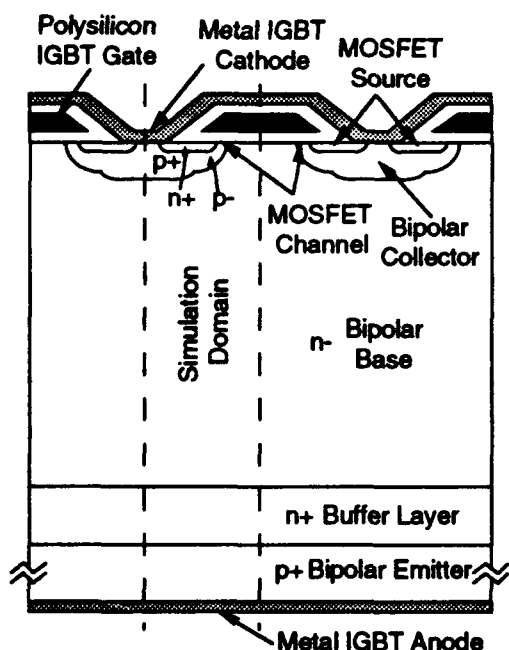


Fig 1. IGBT structure

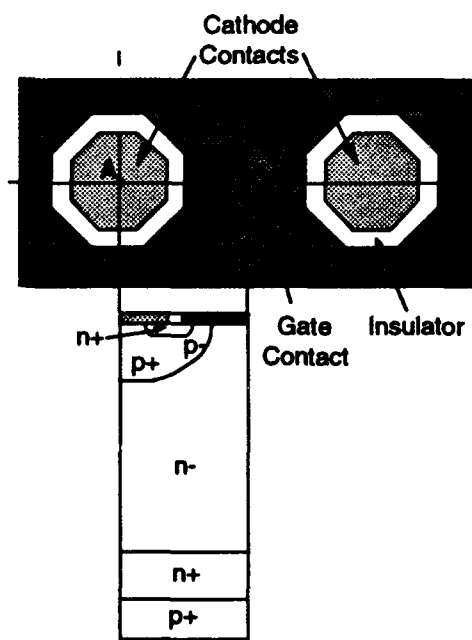


Fig 2. IGBT cell structure and simulation domain

The device behaviour is strongly affected by the cell geometry, MOSFET channel length, actual doping distribution and lifetime killing treatments. The traditional empirical approach to the

device design is inadequate and impractical. It is widely recognised that the device design and optimisation should rely on a proper 2D or even 3D numerical simulation [4]. Although many general purpose simulation packages are now available, they are not usually optimised for the simulation of IGBTs and frequently do this job slowly and unreliably.

Here we report on a new power semiconductor device simulator developed to underpin the design and the optimisation of IGBTs. The simulator have been applied to study the behaviour of relatively short channel IGBTs at high temperatures up to 200°C. Such high temperature operating conditions are typical for the application of these devices in aircraft power conversion systems.

II. The Simulation Program

Our steady state IGBT simulator is based on the solution of the Poisson equation and current continuity equations for electrons and holes. Heavy doping effects are included through bandgap and electron affinity variation in a manner similar to those used for the simulation of compound semiconductor devices. A finite difference method is adopted for discretisation. Although it is well known that the global Newton procedure provides better convergence in the case of strongly coupled equations, a modified Gummel-like iterative scheme is used for nonlinear iterations because it provides a simpler way towards the parallelisation of the simulation code which will be the next step of this development. An appropriate logarithmic dumping for the Poisson equation in combination with a bounded change in electron and hole concentrations [5] ensure convergence in the whole dynamic range of applied on-state voltages up to latch-up conditions. In reverse bias mode the logarithmic dumping provides convergence for the Poisson equation up to several thousand volts. A fast Incomplete LU Factorisation Biconjugate Gradient (ILUBCG) solver is employed for the solution of both Poisson and current continuity equations. The solution time of the three nonlinear equations in the Gummel cycle is significantly reduced if only a few ILUBCG steps follow each nonlinear Newton like step.

In order to simulate properly the high temperature IGBT behaviour appropriate semi-empirical expressions are included for all relevant silicon parameters. The temperature dependent mobility model of Nishida and Sah [6], which includes the scattering mechanisms of surface and bulk acoustical and optical-intervalley phonons, bulk ionised impurities, oxide charges, surface roughness and dipoles or neutral surface states, was modified to account for the carrier-carrier scattering on high injection cases. The recombination terms consist of two components: a Shockley-Read-Hall term and an Auger recombination term. Although slight temperature dependence of the electron and hole lifetimes and Auger coefficient may be expected, due to the lack of relevant data we assume in our analysis that these parameters are temperature independent.

Fig.2. shows an octagonal IGBT cell structure with our solution domain which is a cross section along the line A-A' and which takes into account the symmetry between two neighbouring cells. It is clear however that the 2D simulation is quite a rough approximation to the actual 3D nature of this device and cannot represent cell interaction totally accurately, for example the influence of the distributed parasitic vertical JFET. It is planned to develop a 3D device modeling program implemented on a Parsytec parallel transputer system in order to investigate the effects due to the inherent 3D nature of the device. Additional uncertainty in the results of the simulation is introduced by the approximation for the lateral distribution of the impurities in the MOSFET channel. For a given vertical doping profile (measured using spreading resistance method), experimental results have shown that different assumptions for the lateral impurity distribution can have a marked effect on the subthreshold MOSFET characteristics and the threshold voltage. It is therefore of vital importance when modeling an actual device to have an accurate 2D and in future 3D description of the doping profile, particularly of the channel region, if the results are to give a reliable model of the device characteristics.

III Results

Three typical I_D - V_D curves calculated at room temperature, 100°C and 200°C are given in Fig.3 for $V_G=15V$ which is the expected operating gate voltage for the device. Fig.4. shows a plot of I_D vs V_G for room temperature and 200°C at a drain voltage $V_D=2V$. Both a logarithmic plot (solid lines, left axis) and a linear plot (dotted lines, right axis) are shown on the same graph. The potential, electron and hole distributions ($V_G=15V$, $V_D=2V$) for the top region of the device (to a depth of 15 μm) at room temperature and at 200°C are presented in Fig.5(a,b,c) and Fig.6(a,b,c) respectively. Equipotentials are at -0.15V intervals while equi-concentration contours show the power of ten of the concentration, i.e. contour no.16 represents $10^{16}cm^{-3}$

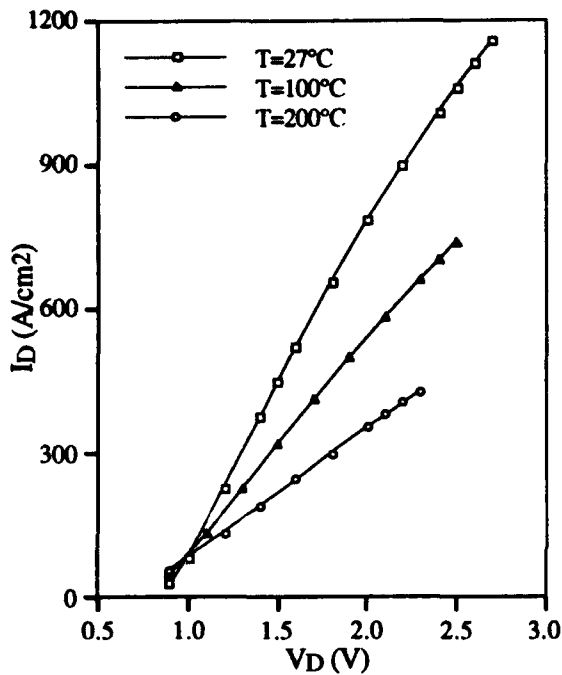


Fig.3. I_D - V_D characteristics at different temperatures ($V_G=15V$)

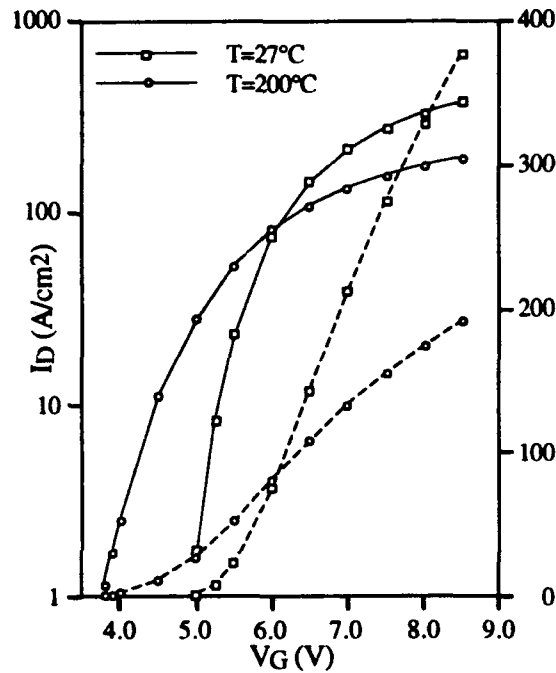


Fig.4. I_D - V_G characteristics at different temperatures ($V_D=2V$)

The high temperature has two main effects on the IGBT characteristics. Firstly the on-state current is reduced by over 50% as the temperature increases to 200°C as can be seen clearly from the I_D - V_D plot and also from the I_D - V_D plot for voltages above the threshold voltage. This means a significant increase in the on-state resistance and on-state power losses. The overall current decrease at 200°C is mainly due to the mobility reduction that decreases the MOSFET current and at the same time increases the resistance of the drift region. The larger voltage drop across the drift region at 200°C, as a result of higher resistance, is clearly seen in Fig.6(a). As can be seen from the comparison of the electron concentrations in Fig.5(b) and Fig.6(b) the carrier concentrations in the drift region in conductivity modulation mode are almost the same (approximately $10^{16} cm^{-3}$).

The I_D - V_G characteristic reflects the MOSFET action of the IGBT gate and as such exhibits the kind of temperature dependence that is expected for a MOSFET [7,8,9]. This is a reduction in the current for voltages above the threshold associated with the drop in mobility, the lowering of the threshold voltage, and a decrease in the sub-threshold slope.

The second important effect at elevated temperature is the decrease in the static latch-up current associated with the turn on of the parasitic thyristor, which exists in the device structure [10]. In

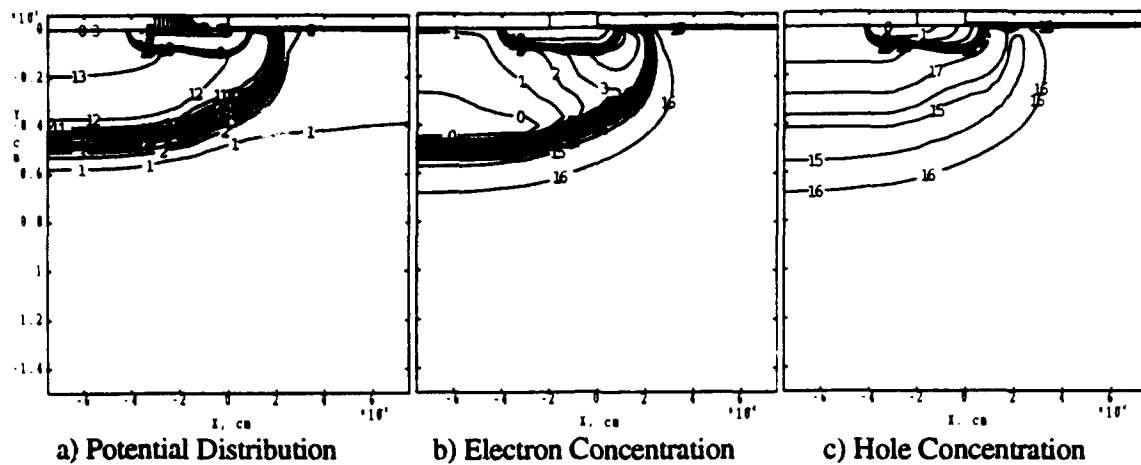


Fig. 5 Potential and carrier concentration distributions at room temperature ($V_G=15V$, $V_D=2V$)

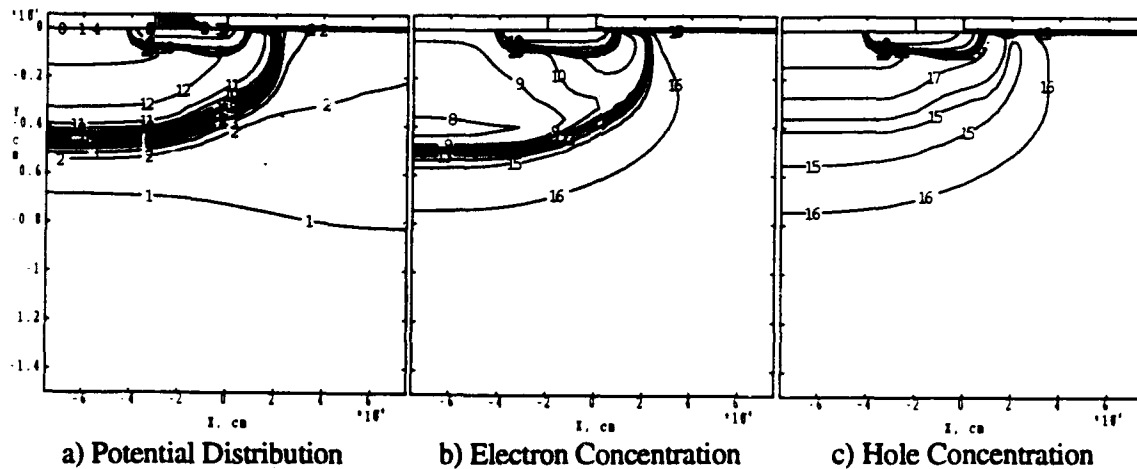


Fig. 6 Potential and carrier concentration distributions at 200°C ($V_G=15V$, $V_D=2V$)

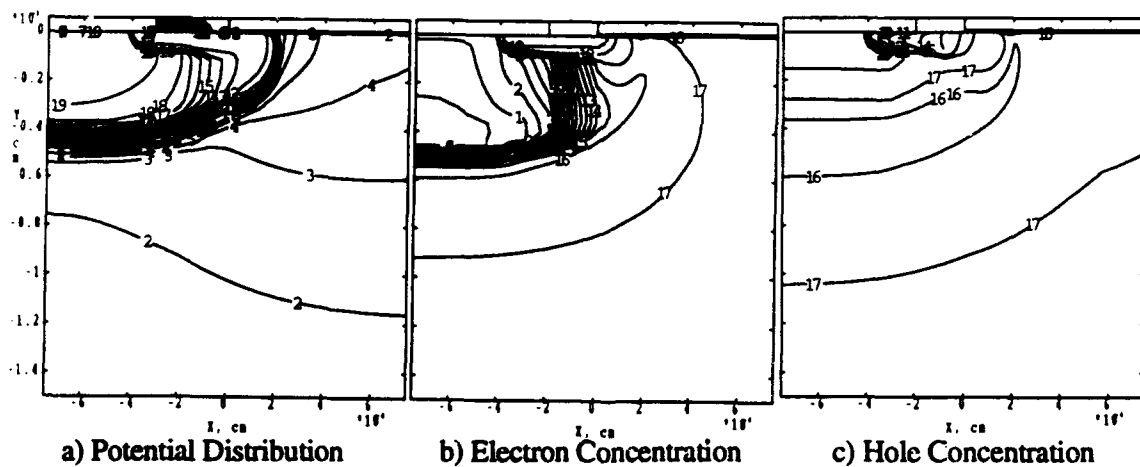


Fig. 7 Potential and carrier concentration distributions at latch-up ($V_G=15V$, $V_D=3V$, $T=27^\circ C$)

the p -region the hole current flows around the n^+ -region and due to the resistance of the p -region a voltage drop is produced. If the current is large enough then this potential difference will be sufficient to forward bias the n^+ - p junction at the channel end injecting electrons from the n^+ -region into the p -region, turning on the parasitic thyristor and latching up the device. At $V_G=15V$ the static latch-up occurs at a current density of approximately $I_D=1150A/cm^2$ at room temperature and at $I_D=440A/cm^2$ at $200^\circ C$. The reduction in the static latch-up current is due to the reduction in the built-in potential of the n^+ - p junction and to the increase in the resistance of the p -region. The potential, electron and hole distribution at latch-up are give in Fig.7(a,b,c). It is interesting to note that the latch-up occurs close to the channel, near the corner of the n^+ -region.

IV. Conclusions

A program has been developed which is suitable for the simulation of IGBTs and which includes the temperature dependence of semiconductor parameters. The study of the high temperature IGBT's operation has shown that increasing the operating temperature of the device results in a significant reduction in on-state drain current for an applied drain voltage. The forward voltage drop across the drift region of the device also increases, as a result of increased resistivity, which leads to an increase in on-state losses. However this is accompanied by the lowering of the threshold voltage. The elevated temperature also reduces the maximum available drain current before latch-up of the device occurs.

Acknowledgements

This work was supported by the United Kingdom Science and Engineering Research Council under grant GR/H23085.

References

- [1] B.J. Baliga, M.S. Adler, R.P. Love, P.V. Gray, and N.D. Zommer, "The Insulated Gate Transistor: A New Three-Terminal MOS-Controlled Bipolar Power Device", IEEE Trans. Electron Dev., Vol. ED-31, No.6, pp.821-828, 1984
- [2] B.J. Baliga, H.R. Chang, T.P. Chow, and S. Al-Marayati, "New Cell Designs for Improved IGBT Safe-Operating Area", IEDM Tech. Dig., pp.809-812, 1988
- [3] B.J. Baliga, "Modern Power Devices", John Wiley & Sons, New York, 1987
- [4] N. Iwamuro, A. Okamoto, S. Tagami and H. Moyoyama, "Numerical-Analysis of Short-Circuit Safe Operating Area for P-Channel and N-Channel IGBTs", IEEE Trans. Electron Dev., Vol.38, No.2, pp.303-309, 1991
- [5] K. Hwang, D.H. Navon, T-W. Tang, and M.A. Asman, "Improved Convergence of Numerical Device Simulation Iterative Algorithm", IEEE Trans. Electron Dev., Vol. ED-32, No.6, pp.1143-1145, 1985
- [6] T. Nishida, C-T. Sah, "A Physically Based Mobility Model for MOSFET Numerical-Simulation", IEEE Trans. Electron Dev., Vol. ED-34, No.2, pp.310-320, 1987
- [7] Y.P. Tsividis, "Operation and Modeling of the MOS Transistor", McGraw-Hill, Singapore, 1988, pp.148-150
- [8] R. Wang, J. Dunkley, T.A. DeMassa, L.F. Jelsma, "Threshold Voltage Variations with Temperature in MOS Transistors", IEEE Trans. Electron Dev. Vol. ED-18, No.6, pp.386-388, June 1971
- [9] L. Valdez, A.S. Grove, "Temperature Dependence of MOS Transistor Characteristics Below Saturation", IEEE Trans. Electron Dev., Vol. ED-13, No.12, pp.863-866, 1966
- [10] N. Mohan, T.M. Undeland, W.P. Robbins, "Power Electronics: Converters, Applications and Design," John Wiley & Sons, New York, 1989

Deterministic BJT Modeling by Self-Consistent Solution to the Boltzmann, Poisson and Hole-Continuity Equations

Hongchin Lin*, Neil Goldsman and I.D. Mayergoyz

*Advanced Micro Devices, Sunnyvale, CA 94088, U.S.A.

Department of Electrical Engineering,
University of Maryland, College Park, MD 20742, U.S.A.

Abstract

A new comprehensive and efficient bipolar junction transistor (BJT) device model is presented. This model self-consistently solves the Boltzmann transport equation (BTE) for electrons, the current-continuity equation for holes and the Poisson equation. The calculations provide almost the same information as similar Monte Carlo simulations, but require only 1% of the CPU time. In addition, a new discretization has been employed which facilitates convergence.

I. Introduction

In the past, device modeling by direct solution of the Boltzmann transport equation (BTE) was usually considered too difficult to be achieved. The difficulties were mainly due to dimensionality problems (the steady-state BTE is a 6-dimensional equation), and problems evaluating the complicated collision integral.

Recently, however, the Legendre polynomial (LP) technique has been demonstrated to provide fast and accurate solutions to the BTE[1,2,3]. Use of the LP technique allows one to overcome problems of dimensionality, and also facilitates evaluation of the collision integrals. However, until now the method was used only for very simple, largely unrealistic test structures[2], or only as a post-processor to hydrodynamic simulations[3]. Here, we overcome these limitations and adapt the new LP technique to actually model BJT's.

II. General Approach

In the present work we self-consistently solve the electron BTE, the Poisson, and hole-continuity equations to obtain the electron momentum distribution function for an entire prototype BJT. We first transform the BTE into a tractable expression using a first order LP expansion: this reduces the dimensionality of the system, and allows us to integrate the collision terms analytically.

To numerically solve the system, we first discretize the BTE using a new Scharfetter-Gummel-like algorithm which employs special state variables. After discretizing the Poisson and hole-continuity equations, the overall nonlinear system is then solved using a Gummel-type iteration scheme. This algorithm provides nearly exponential convergence to efficiently obtain the distribution function for the entire device without the statistical noise which is characteristic of Monte Carlo methods.

III. The Boltzmann-Poisson-Continuity System

The device model consists of the Poisson equation (1), the space-dependent BTE for electrons (2), and the current continuity-equation for holes (3). The collision integrals in the BTE account for the effects of acoustic and intervalley phonons, as well as impact ionization scattering. These scattering cross-sections, as well as nonparabolic conduction band-structure values, are identical to those employed in Monte Carlo calculations[4].

$$\nabla_{\mathbf{r}}^2 \phi(\mathbf{r}) = \frac{e}{\epsilon_s} [n(\mathbf{r}) - p(\mathbf{r}) + N_A(\mathbf{r}) - N_D(\mathbf{r})] \quad (1)$$

$$\frac{1}{\hbar} \nabla_{\mathbf{k}} \epsilon \cdot \nabla_{\mathbf{r}} f(\mathbf{k}, \mathbf{r}) + \frac{e}{\hbar} \nabla_{\mathbf{r}} \phi(\mathbf{r}) \cdot \nabla_{\mathbf{k}} f(\mathbf{k}, \mathbf{r}) = \left[\frac{\partial f(\mathbf{k}, \mathbf{r})}{\partial t} \right]_{ac} + \left[\frac{\partial f(\mathbf{k}, \mathbf{r})}{\partial t} \right]_{iv} + \left[\frac{\partial f(\mathbf{k}, \mathbf{r})}{\partial t} \right]_{ii} \quad (2)$$

$$\nabla_{\mathbf{r}} \cdot [\mu_p p(\mathbf{r}) \nabla_{\mathbf{r}} \phi(\mathbf{r}) + \mu_p V_i \nabla_{\mathbf{r}} p(\mathbf{r})] = R(\phi, n, p) \quad (3)$$

where $n(\mathbf{r}) = \frac{1}{4\pi^3} \int f(\mathbf{k}, \mathbf{r}) d\mathbf{k}$ is the electron concentration; $p(\mathbf{r})$ is the hole concentration; $\phi(\mathbf{r})$ is the potential; $N_D(\mathbf{r})$ and $N_A(\mathbf{r})$ are the doping concentration for donors and acceptors; $f(\mathbf{k}, \mathbf{r})$ is the distribution function; R is the recombination rate, including impact ionization and Shockley-Read-Hall (SRH) recombination; $V_i = K_B T / e$; the subscripts *ac*, *iv*, *ii* correspond to acoustic phonons, intervalley phonons, impact ionization, respectively.

IV. Method of Solution:

1. BTE Formulation: Legendre Expansion

We first express the distribution function in terms of Legendre polynomial basis functions:

$$f(\mathbf{k}, \mathbf{r}) = f(\mathbf{k}, x) = f_0(\epsilon, x) + kg(\epsilon, x) \cos \theta \quad (4)$$

where θ is the angle between $\nabla_{\mathbf{r}} \phi(\mathbf{r})$ and \mathbf{k} ; $f_0(\epsilon, x)$ and $kg(\epsilon, x)$ represent the coefficients of the symmetrical and the anti-symmetrical parts of the distribution function respectively.

In solving the BTE, our objective will be to determine the unknown coefficients f_0 and g . To determine these coefficients, we next substitute the Legendre expansion into the BTE, and use symmetry to obtain 2 equations for the 2 unknowns f_0 and g .

2. Numerical Solution of BTE: A Scharfetter-Gummel-Like Approach

At this point, the typical approach would be to discretize and try to solve the resulting equations directly. However, this direct approach would lead to a discrete matrix which is ill-conditioned, would not readily account for the exponential variation in the distribution function, and would inhibit obtaining a solution to the overall Boltzmann-Poisson-Continuity system.

To overcome numerical problems, and routinely solve the coupled system, we developed a Scharfetter-Gummel-like discretization scheme to resolve the exponential behavior of the distribution function. This scheme enhances the diagonal elements of the discrete coefficient matrix, and helps to numerically account for the rapid variation in the distribution function. With this approach, $f_0(\epsilon, x)$ is expressed as the Slotboom-like form

$f_0(\epsilon, x) = n(x)q(\epsilon, x)$; and $n(x)$ is given as $n_i u(x) \exp(\phi(x)/V_i)$. We then substitute these new variables into the symmetrical and antisymmetrical equations obtained from the original BTE, and discretize using finite differences. We next analytically integrate the rapidly varying part of the difference equations between grid points. This allows us to account for the exponential variation of the distribution function analytically, thereby alleviating the computer of much of the burden. The discretization is then completed to yield a matrix equation with significantly enhanced diagonal terms. This discrete version of the BTE is then solved using SOR-type iterations in the real-space domain and sparse-matrix Gaussian elimination in energy-space.

3. Self-Consistently Solving the Coupled System

With a robust method for solving the BTE developed, the entire coupled Boltzmann-Poisson-Continuity nonlinear system is solved using a Gummel-type iteration scheme. The Poisson equation is directly discretized and solved with sparse matrix algebra. The hole-current-continuity equation is discretized with the standard Scharfetter-Gummel approach, and solved directly with sparse matrix Gaussian elimination. To facilitate convergence of the overall scheme, special damping and weighting factors have been developed which help to guide each Gummel iteration toward the proper solution. The flow chart of this numerical procedure is shown in Fig. 1.

V. Results

Simulation results of a submicron $n^+/p/n^-/n^+$ BJT are shown in Figs. 2 to 8. In Fig. 2 the prototype BJT structure is shown. Fig. 3 shows the distribution function for the entire device. In Figs. 4 and 5 we show calculated values for average velocity, carrier concentration, average energy and electric field. The figures also show that good agreement with MC calculations, which employ the same transport model, was obtained. It is worth noting that velocity overshoot, which is characteristic of non-equilibrium electron transport, is observed near the p/n^- junction. To demonstrate the robustness of the algorithm, in Figs. 6 to 8 we show results of average energy and velocity, as well as ionization coefficients, calculated for a large range of applied biases.

VI. Conclusion

We have developed an accurate and stable approach to BJT simulation by the direct, self-consistent solution to the Poisson, hole-continuity and electron-Boltzmann equations. The method calculates the distribution function for the entire device. Furthermore, the method uses less than 1/100 the CPU time required by similar MC calculations.

References

- [1] H. Lin, N. Goldsman and I.D. Mayergoyz, *Solid-State Electronics*, vol. 35, p. 33 (1992).
- [2] H. Lin, N. Goldsman and I.D. Mayergoyz, *Solid-State Electronics*, vol. 35, p. 769 (1992).
- [3] A. Gnudi, D. Ventura, G. Baccarani, *Proceedings SISDEP '91*, p. 205.
- [4] C. Jacoboni and L. Reggiani, *Reviews of Modern Physics*, 55, p.645 (1983).

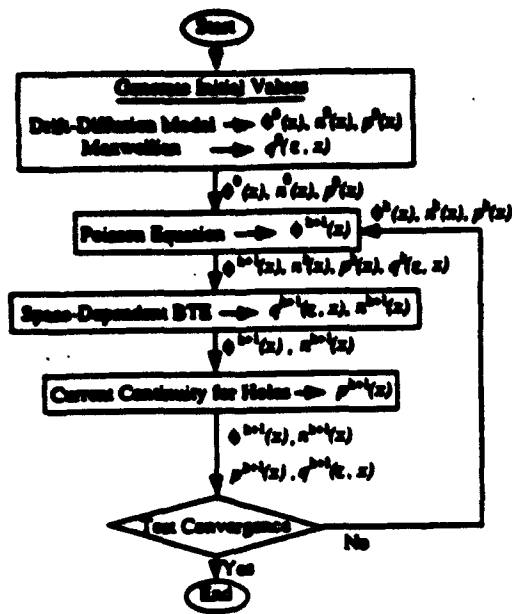


Fig. 1. The flow chart illustrates the Gummel algorithm used in this work.

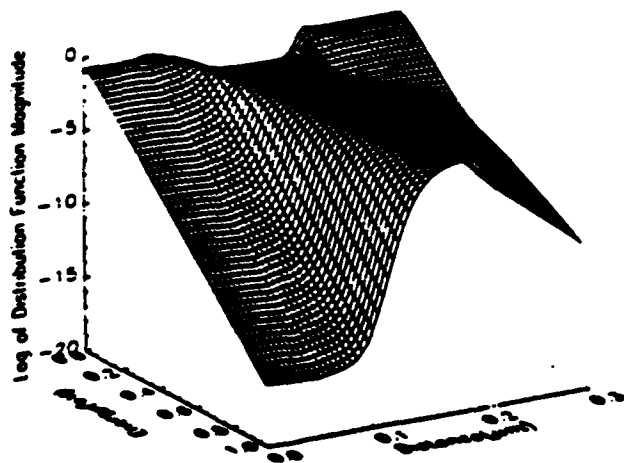


Fig. 3. The energy distribution function as a function of position along the BJT.

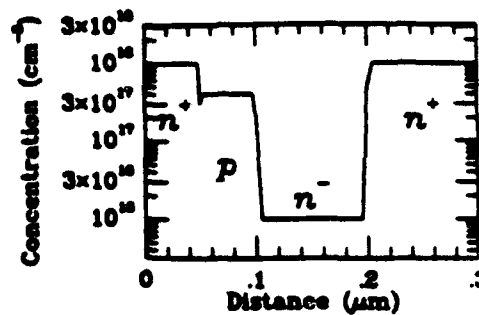


Fig. 2. The prototype BJT modeled by direct solution of the Boltzmann-Poisson-Continuity system.

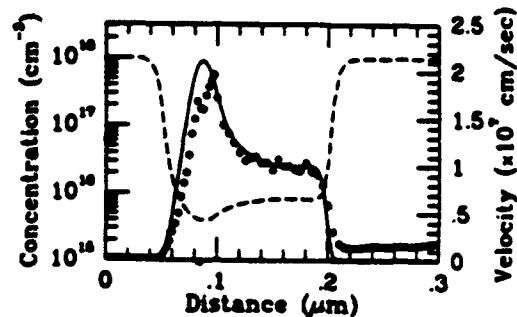


Fig. 4. Electron concentration (dash line) and average velocity (solid line) calculated by the new method. Average velocity from MC calculations (open circles) is also plotted.

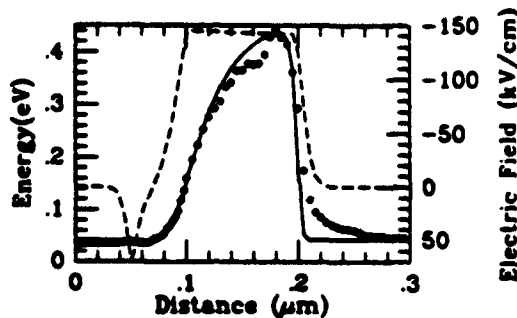


Fig. 5. Electric field (dash line) and average energy (solid line) calculated with this direct method. MC calculations are also plotted (open circles).

Fig. 6. Average electron energy as a function of position along the BJT. The simulations easily converge for a wide range of applied biases using the Scharfetter Gummel-like discretization of the Boltzmann-Poisson-Continuity system.

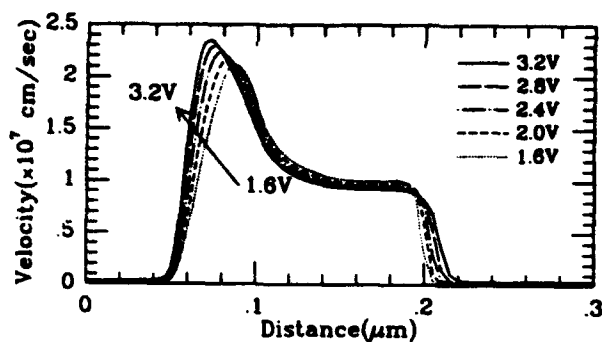
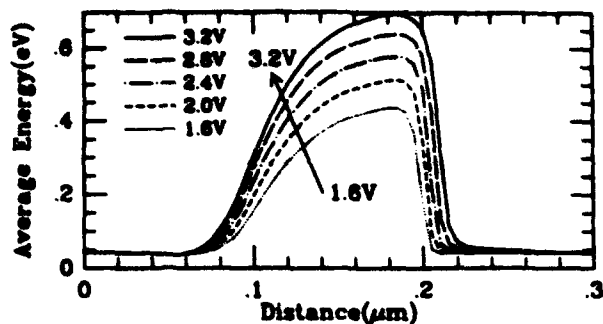
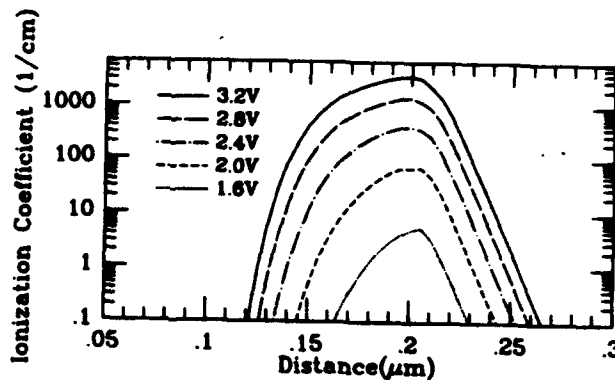


Fig. 7. Average electron velocity as a function of position calculated by the Boltzmann-Poisson-Continuity system for a wide range of applied biases. Velocity overshoot is clearly predicted by the method.

Fig. 8. Values for impact ionization coefficient calculated for a wide range of applied bias voltages. It is clear that impact ionization is maximum in the high-field collector region.



A Self-Consistent Charge-Control Model for HEMTs Incorporating Deep Level Effects

Ranjit Singh and Christopher M. Snowden

**Microwave and Terahertz Technology Group,
Department of Electrical and Electronic Engineering,
University of Leeds, United Kingdom**

Abstract

A new self-consistent charge-control model which includes models for realistic trapping processes is described for a high electron mobility transistor (HEMT). This generalised HEMT model takes into account the effects of deep levels both in semi-insulating GaAs and doped AlGaAs layers. It is shown that conventional charge-control models are insufficient to describe HEMT operation near pinch-off and in the normally-on mode. The model is used to investigate the effects of Fermi-level pinning at the semi-insulating interface on the device performance and is shown to lead to less injection of electrons into substrate.

I. Introduction

An understanding of the trapping mechanism in HEMTs is very important for developing accurate models for HEMT characterisation. Trapping of electrons in doped AlGaAs layer and semi-insulating (SI) GaAs substrate is not only responsible for different anomalies observed at cryogenic temperatures but limit the device performance at 300K as well. Examples include the collapse of I-V characteristics, decrease in both maximum transconductance and in the gate voltage swing, shift in pinch-off voltage, slow transients in switching, and generation-recombination noise. These effects become more pronounced in very short channel HEMTs for millimeter-wave applications. Although trapping phenomena are very important in determining the characteristics of HEMTs and there has been significant amount of research directed towards understanding the basic physics of deep levels in doped AlGaAs layers[1]-[2] and SI substrate [3]-[4], there has been little theoretical effort devoted to incorporating trapping mechanisms in the device models. The influence of Fermi-level pinning at the GaAs substrate on HEMT operation has been discussed by Krantz et al [5] using analytical expressions. Shawki et al [6] have included in their two-dimensional model the trapping process associated with the doped AlGaAs layer by assuming that all donors are electrically active as DX centers. They demonstrated that the DC transconductance can be lowered as much as 60% due to electrons trapped in the doped AlGaAs layer. Japanese work [7] has confirmed that DX centers are also present in n-type GaAs and become the ground state when pressure exceeds 20 K bar. This result provides the strong evidence that DX properties are associated with isolated donors. The measurement work done at IBM Thomas J. Watson Research Center [2] has further shown that there is a strong variation of trapping

kinetics with change in alloy composition of $\text{Al}_x\text{Ga}_{1-x}\text{As}$. These results strongly suggest that trapping models should include more than one donor level.

In our recent paper [8], we have presented a more realistic three-level trapping model for doped AlGaAs layer. In this paper we present a new self-consistent charge-control model which includes realistic trapping processes in both the doped AlGaAs and SI substrate. The model provides a natural explanation of the published experimental results related to trapping mechanisms and demonstrates the role of deep level states on the operation of single-, multi-channel, inverted and pseudomorphic HEMTs.

II. The Charge-Control Model

In this section a self-consistent charge-control model incorporating deep level effects is developed by modifying the Poisson equation to include trapping effects. Based on the results of a self-consistent quantum mechanical model, which shows that most of the electron reside in first three quantum energy levels, and on the study of Yoshida [9] that a classical model with Fermi-statistics predicts the device performance with good accuracy, a fast and accurate charge-control model is developed. The present model is based on self-consistently solving Poisson and Schrödinger equations for a maximum of up to three quantum energy levels and applying a classical model with Fermi-statistics for the remaining electrons. The self-consistent quantum mechanical model has been described in the literature [9]-[10] and only the broad principle will be described here. The electrostatic potential is related to charge-distribution by Poisson equation

$$\frac{\partial}{\partial x} \left(e \frac{d}{dx} V(x) \right) + \rho(x) = 0 \quad (1)$$

where ρ is the net local charge density, and the envelope wavefunction Ψ_i for the i th sub-band satisfies the schrodinger equation

$$\frac{\hbar^2}{2} \left[\frac{d}{dx} \left(\frac{1}{m^*} \frac{d\Psi_i}{dx} \right) \right] + [E_i - V_i(x)] \Psi_i(x) = 0 \quad (2)$$

where m^* is the effective mass, E_i is the eigen energy for the i th sub-band and V_i is the potential including the local exchange correlation coefficient. The Fermi level is assumed to be constant throughout the semiconducting layers in equilibrium and the self-consistent solution of equations 1 and 2 gives the accurate band profile, sub-band energies and Fermi level.

A. The Substrate Trap Model

A high density of both donor- and acceptor-like traps exists in the SI substrate. It is generally accepted that there are up to 4 traps which are important in determining the properties of SI

GaAs [3]-[4] and only these traps are included in the present charge-control model. The first of these is EL2 level deep donor which resides at an energy level 0.8eV below the conduction band edge in concentrations ranging between 10^{20} and 10^{23} m^{-3} . The shallow donor and acceptor traps are generally present in concentration of about 10^{23} m^{-3} . The shallow donor is attributed to silicon introduced from the walls of reactor vessel and shallow acceptor is associated to carbon. Finally, a deep acceptor level is present at an energy level about 0.8 eV above the valence band edge. The local net charge density in SI substrate is given by

$$\rho = q [N_D + N_{DTn} + N_{DTd} - N_{ATn} - N_{ATd} - n + p] \quad (3)$$

where N_D is the doping density, N_{DTn} , N_{DTd} are the net densities of positive trapped charge in the shallow and deep donor traps and N_{ATn} , N_{ATd} are the net densities of negative trapped charge in shallow and deep acceptors traps respectively.

B. The Doped Channel Trap Model

The doped AlGaAs layer introduces several undesirable effects for HEMT operation. They are mainly related to large number of deep states which become dominant for Al mole fractions ≥ 0.22 . The doped layer trapping mechanism is incorporated in the present charge-control model based on simple three donor-level model which is described in detail in our recent paper [8]. However, as we will refer to it in this paper, a brief description of it is included here for the sake of completeness. The doped donor atoms can occupy either of the three donor levels, shallow, deep, or DX. The total donor doping density in our model

is given by the sum of the densities of the three donor levels. This model is consistent with the models treated theoretically by Morgan [1] and also accounts for experimental results obtained by Mooney et al [2]. The local net charge density in AlGaAs layer is given by

$$\rho = q [N_D - N_{ST} - N_{dT} - N_{DXT} - n + p] \quad (4)$$

where N_{ST} , N_{dT} , N_{DXT} are respectively the net densities of electrons trapped in the shallow, deep and DX levels.

C. The Trap Filling Factor

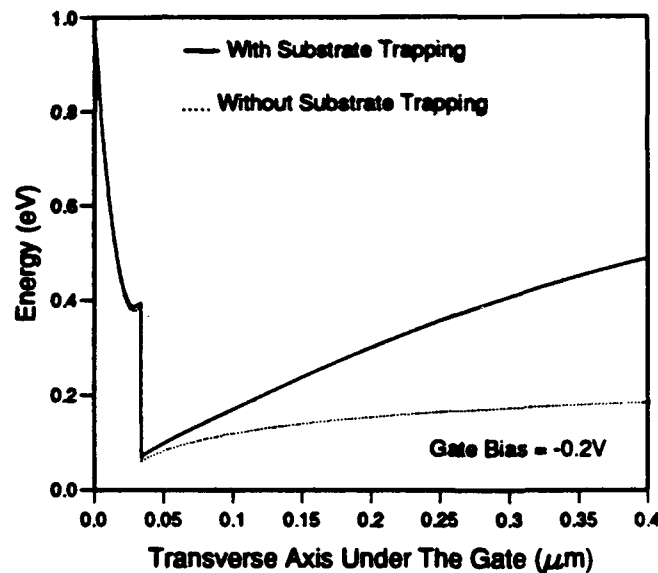
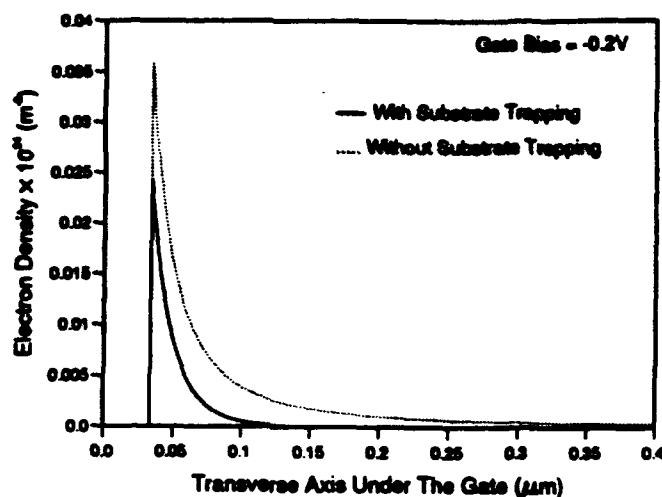


Figure 1. Conduction band energy of a single channel HEMT along the transverse axis under the gate.

In HEMTs, the density of mobile carriers depends on the electron trapping in the trap levels. The filling of these trap levels is calculated based on a Shockley-Read-Hall model [11]-[12] for recombination through a single level. The donor traps are positive when empty (i.e. when occupied by holes) and neutral when filled (i.e. when occupied by electrons), and acceptor traps are negative when filled and neutral when empty.



III. Results and Discussions

Figures 1 and 2 respectively show the single-channel HEMT's conduction band energy and electron density along the transverse axis under the gate, with and without the SI substrate trapping effect. The simulation results have shown that trapping in the SI substrate have a significant effect near pinch-off. The pinning of the Fermi level at the SI interface makes the conduction band more abrupt and an interface depletion region is formed. This results in decreased electron density near the SI interface due to narrowing of the channel. At the same time this leads to greater confinement of channel electrons and less substrate electron injection. Due to improved carrier confinement with the SI substrate, the device has good pinch-off characteristics.

The simulation results have shown that doped layer trapping effects become more important as the gate-source voltage is made more positive. This is due to the increase in parallel current in doped AlGaAs layer and saturation of two-dimensional electron gas which no longer responds to variation in gate voltages. This demonstrates the decrease in maximum transconductance and collapse of IV characteristics at high gate voltages. Figure 3 shows the charge-control results of a two-channel AlGaAs/GaAs HEMT and demonstrates that doped layer trapping effect is dominant for all the gate-biases. In multi-

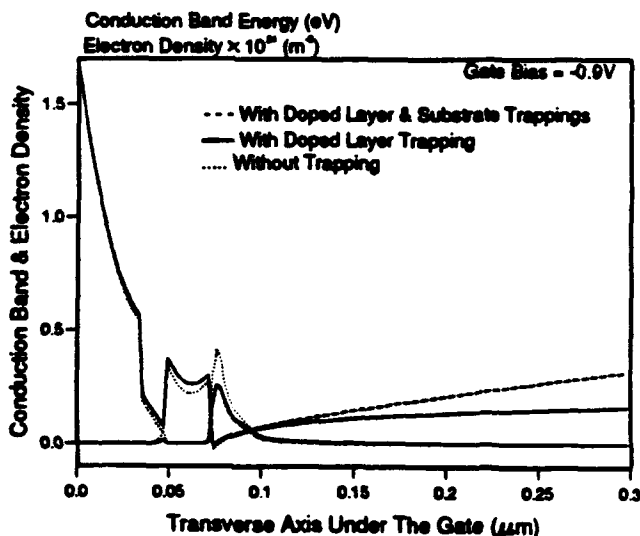


Figure 3. Conduction band energy and electron density of a two channel HEMT along the transverse axis under the gate.

channel HEMTs, the effect of trapping is more pronounced in the second doped layer because parallel current in it flows for almost all the gate-bias voltages. In pseudomorphic HEMTs, the effect of the SI substrate is less pronounced.

IV. Conclusions

We have developed a new self-consistent charge-control model for HEMTs that includes trapping effects in the SI GaAs substrate and doped AlGaAs layer and provides a physical understanding of short channel HEMT operation. It is expected that the present charge-control model will be useful in developing more realistic and accurate models for HEMTs.

References

- [1] T. N. Morgan, "Theory of the DX center in $\text{Al}_x\text{Ga}_{1-x}\text{As}$ and GaAs crystals", *Phys. Rev. B*, vol. 34, pp. 2664-2669, Aug. 1986.
- [2] P. M. Mooney, N. S. Caswell and S. L. Wright, "The capture barrier of DX center in Si doped AlGaAs," *J. Appl. Phys.*, vol.62, pp.4786-4797, Dec.1987.
- [3] E. J. Johnson, J. A. Kafalas, and R. W. Davies, "The role of deep level centers and compensation in producing semi-insulating GaAs," *J. Appl. Phys.*, vol. 54, no.1, pp. 204-207, Jan. 1983.
- [4] P. F. Lindquist and W. M. Ford, "Semi-insulating GaAs substrates," in *GaAs FET Principles and Technology*, J. V. Dilozenzo and D. D. Khandelwal, Eds. Deadham, MA: Artech. 1982, pp.1-60.
- [5] R. J. Krantz, D. C. Mayer, and W. L. Bloss, "The influence of Fermi-level pinning at the GaAs substrate on HEMT threshold voltage," *Solid-State Electronics*, vol. 33, no. 9, pp. 1189-1195, 1990.
- [6] T. Shawki, G. Salmer and O. El-Sayed, "MODFET 2-D hydrodynamic energy modelling: optimization of subquarter-micron-gate structures", *IEEE Trans. Electron Devices*, vol.ED-37, pp. 21-29, Jan. 1990.
- [7] M. Tachikawa, T. Fujisawa, H. Kukimoto, A. Shibata, G.Oomi, and S.Minomura, *Japan J. Appl. Phys.*, vol. 24, pp. L893, 1985.
- [8] R. Singh and C. M. Snowden, "Investigation of trapping effect on the operation of HEMT", *Proceedings GaAs Simulation Group VII Meeting*, Harrogate, April 1993.
- [9] J. Yoshida, "Classical versus quantum mechanical calculation of the electron distribution at the n-AlGaAs/GaAs heterointerface," *IEEE Trans. Electron Devices*, vol.ED-33, pp.154-156, Jan 1986.
- [10] B. Vinter, "Subbands and charge control in a two-dimensional electron gas field-effect transistor," *Appl. Phys. Lett.* 44(3), pp. 307-309, Feb. 1984.
- [11] W. Shockley and W. T. Read, "Statistics of recombination of holes and electrons," *Phys. Rev.*, vol.87, pp. 835-842,1952.
- [12] R. N. Hall, "Electron hole recombination in germanium," *Phys. Rev.*, vol. 87, pp. 387, 1952.

Transport Coefficients For A GaAs Hydrodynamic Model Extracted From Inhomogeneous Monte Carlo Calculations

Mei-Kei leong and Ting-wei Tang

Department of Electrical and Computer Engineering
University of Massachusetts, Amherst MA 01003, USA

Abstract

The validity of energy-dependent relaxation times often used in a GaAs hydrodynamic model has been carefully examined using the self-consistent Monte Carlo simulation. We have found that those transport coefficients associated with the intervalley transfer from the lower to the upper valley are not single-valued functions of the averaged electron energy in the valley. If, instead, the valley population ratio or the average energy weighted by the valley population ratio is used a substantial improvement in accuracy can be achieved.

I. Introduction

Conventionally, macroscopic (averaged) relaxation times such as τ_n , τ_p and τ_w appearing in the hydrodynamic (HD) transport equations are assumed energy-dependent and determined by performing Monte Carlo (MC) calculations under steady state and homogeneous field conditions. These expressions are often extended to the case of inhomogeneous fields without any justification. Sandborn et al. [1], using the MC simulation, found that both the energy and momentum relaxation times in an equivalent single-valley model under the transient condition differ very much from the steady state and homogeneous field values. Yamada [2] also observed the discrepancy in the relaxation times between the homogeneous and inhomogeneous field conditions. He suggested that the relaxation times should depend not only on the energy but also on the valley population. In this study, instead of a single-valley model, a three-valley HD model for GaAs has been developed. The transport coefficients appearing in the HD model for homogeneous and inhomogeneous field conditions are evaluated by a single particle MC simulation program and a multi-particle self-consistent MC simulation program, respectively. A one dimensional $N^+ - N - N^+$ GaAs ballistic diode was used as a test device. This approach allows us to rigorously examine the validity of the energy dependence of each relaxation time as well as provides us valuable information for a more appropriate description of the relaxation times.

II. Moments of the Boltzmann Transport Equation

The HD transport equations can be obtained by taking various moments of the Boltzmann transport equation (BTE) [3]. Extending the work of [4], [5] to a multi-valley system, we obtain the following steady-state conservation equations for the i th valley:

$$\nabla \cdot (n_i \vec{V}_i) = \left[-\frac{n_i}{\tau_{nij}} - \frac{n_i}{\tau_{nih}} + \frac{n_j}{\tau_{nji}} + \frac{n_h}{\tau_{nhi}} \right], \quad (1)$$

$$\frac{1}{n_i} \nabla \cdot (n_i \hat{U}_i) - \vec{F} = -\frac{q\vec{V}_i}{\mu_{ii}} - \frac{q\vec{V}_i}{\mu_{ij}} - \frac{q\vec{V}_i}{\mu_{ih}}, \quad (2)$$

$$\frac{1}{n_i} \nabla \cdot (n_i \vec{S}_i) - \vec{V}_i \cdot \vec{F} = -\frac{W_i - W_0}{\tau_{Wii}} - \frac{W_i}{\tau_{Wij}} - \frac{W_i}{\tau_{Wih}} + \frac{n_j}{n_i} \frac{W_j}{\tau_{aji}} + \frac{n_h}{n_i} \frac{W_h}{\tau_{ahi}}, \quad (3)$$

$$\frac{1}{n_i} \nabla \cdot (n_i \hat{R}_i) - (W_i \hat{J} + \hat{U}_i) \cdot \vec{F} = -\frac{q\vec{S}_i}{\mu_{iii}} - \frac{q\vec{S}_i}{\mu_{oij}} - \frac{q\vec{S}_i}{\mu_{oih}}, \quad (4)$$

where $\vec{V}_i = \langle \vec{v}_i \rangle$, $\hat{U}_i = \langle \vec{v}_i \hbar \vec{k}_i \rangle$, $W_i = \langle \epsilon_i \rangle$, $\vec{S}_i = \langle \vec{v}_i \epsilon_i \rangle$ and $\hat{R}_i = \langle \epsilon_i \vec{v}_i \hbar \vec{k}_i \rangle$. In the conventional HD transport model, the transport coefficients (τ_n, μ, τ_W , etc.) are usually assumed to depend on the average energy, $W(\vec{r})$. This approach ignores the dependence of the transport coefficients on the shape of the distribution function.

III. Self-Consistent Monte Carlo Simulation

To examine the accuracy of energy-dependent transport coefficients, we begin with a rigorous solution of the steady-state BTE by the MC method. The simulation program uses an analytical multi-valley, non-parabolic band. The following types of scattering are taken into account: acoustic phonon scattering, optical phonon scattering, polar optical phonon scattering, ionized impurity scattering, equivalent and nonequivalent intervalley scattering. The various scattering parameters are similar to those used in [6]. In this work, a one-dimensional $N^+ - N - N^+$ GaAs structure with a $0.5 \mu m$ N-region was examined. The doping densities of the three layers were $N_d = 1 \times 10^{17} cm^{-3}$, $1 \times 10^{16} cm^{-3}$ and $1 \times 10^{17} cm^{-3}$, respectively. The applied bias was 2.0 volts. Fig. 1 displays the doping density and electric field profiles within the device as obtained from the self-consistent MC (SCMC) simulation. Fig. 2 shows the Γ -valley velocity and energy profiles.

IV. Results and Discussions

At each position within the device we evaluated the transport coefficients and the average energy in each valley by the SCMC program. We found that these coefficients are generally a function of the local average energy in the valley except for those due to the intervalley transfer from the lower to the upper valley (i.e. $\Gamma \rightarrow L$, $\Gamma \rightarrow X$, and $L \rightarrow X$). Figs. 3, 5 and 7 respectively display the $\Gamma \rightarrow L$ intervalley transfer coefficients, $\tau_{n\Gamma L}$, $\mu_{\Gamma L}$ and $\tau_{W\Gamma L}$ versus the average Γ -valley energy, W_Γ , for both the homogeneous and the inhomogeneous field calculation. The "hysteresis" loops clearly indicate that none of them can be described as a single-valued function of W_Γ . For these cases, the energy-dependent transport coefficients which were obtained from the homogeneous field calculation always underestimate the actual one in the increasing field region and overestimate the same in the decreasing field region. If, instead, W_Γ is weighted by the valley population ratio (i.e., $\frac{n_L}{n_\Gamma} W_\Gamma$) then the hysteresis loop for $\tau_{n\Gamma L}$ can be significantly reduced (see Fig. 4). This is motivated by the fact that the valley population ratio more or less reflects the fraction of electron population which have sufficient energy to transfer from the lower to the upper valley. We also found that using the valley population ratio alone (i.e., $\frac{n_L}{n_\Gamma}$) the hysteresis loop can be considerably reduced for $\mu_{\Gamma L}$ and $\tau_{W\Gamma L}$ as seen in Figs. 6 and 8, respectively.

The result for τ_{eRL} is very similar to τ_{wRL} and that for μ_{eRL} is similar to μ_{RL} . Once the hysteresis loop is reduced, these transport coefficients can now be modelled empirically as single-valued functions of the valley population ratio or the energy weighted by the valley population ratio.

V. Conclusions

A SCMC simulation program was used to examine the conventional assumption of energy-dependent transport coefficients in a multi-valley system. We found that the transport coefficients related to the intervalley transfer from the lower to the upper valley (i.e., $\Gamma \rightarrow L$, $\Gamma \rightarrow X$ and $L \rightarrow X$) are not a single-valued function of the average energy in the valley. A substantial improvement in the accuracy can be achieved if the valley population ratio or the average energy weighted by the valley population ratio is used.

Acknowledgement

The authors acknowledge Prof. D. Navon for carefully reading the manuscript. This work was supported in part by NSF Grant ECS-9003518.

References

- [1] P.A. Sandborn, A. Rao and P. A. Blakey, *IEEE, Trans. Electron Devices*, Vol.36, No.7, pp.1244-1253, Jul., 1989.
- [2] Yoshinori Yamada, *IEICE Trans.*, Vol.E73, No.2, pp.255-259, Feb., 1990.
- [3] Blotekjær, K., *IEEE Trans. Electron Devices*, Vol. ED-17, No.1, pp.38-47, Jan., 1970.
- [4] S-C. Lee and T-w. Tang, *Solid State Elec.*, Vol.35, P.561-569, Apr., 1992.
- [5] T-w. Tang, S. Ramaswamy and J. Nam, to appear in *IEEE Trans. Electron Devices*.
- [6] M.A. Littlejohn, J.R. Hauser, T.H. Glisson, *J. Appl. Phys.*, Vol.48, pp.4587-4590, Nov., 1977.

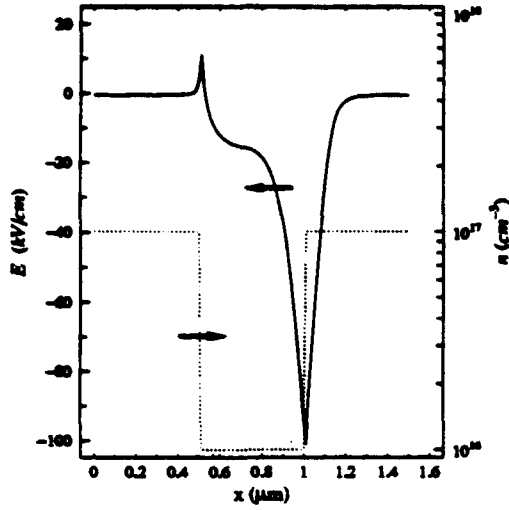


Figure 1: Impurity doping profile and its self-consistent electric field for an abrupt $N^+ - N - N^+$ structure with an applied bias of 2.0 volts.

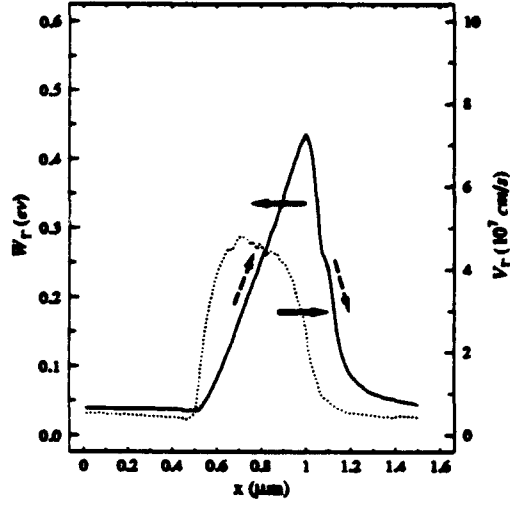


Figure 2: Average electron energy and velocity in the Γ -valley.

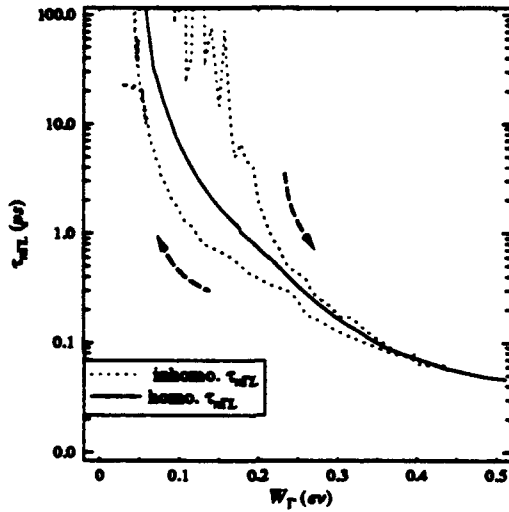


Figure 3: Carrier relaxation time, $\tau_{n\Gamma L}$, versus the average Γ -valley energy, W_Γ .

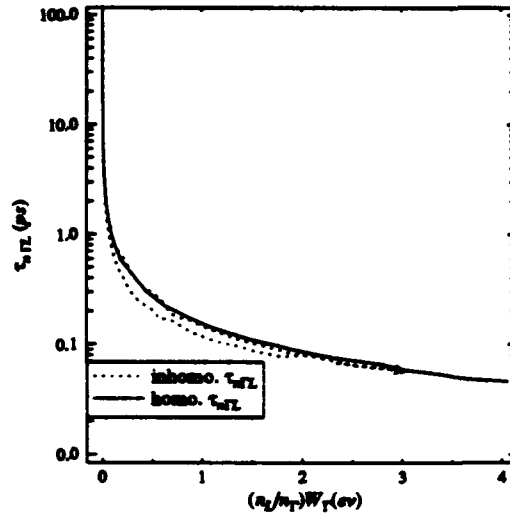


Figure 4: Carrier relaxation time, $\tau_{n\Gamma L}$, versus $(n_0/n_T)W_\Gamma$.

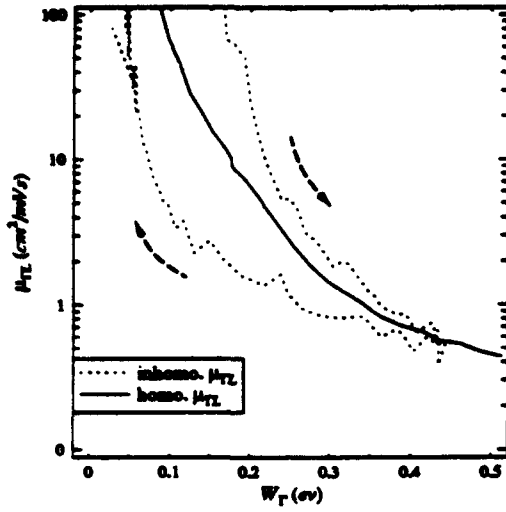


Figure 5: Carrier mobility, μ_{TL} , versus the average Γ -valley energy, W_Γ .

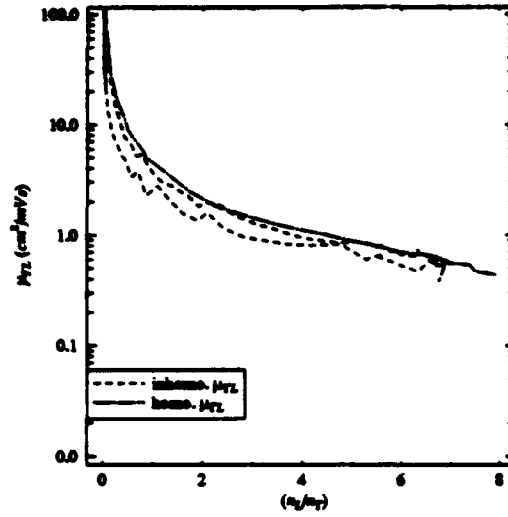


Figure 6: Carrier mobility, μ_{TL} , versus $(\frac{n_L}{n_T})$.

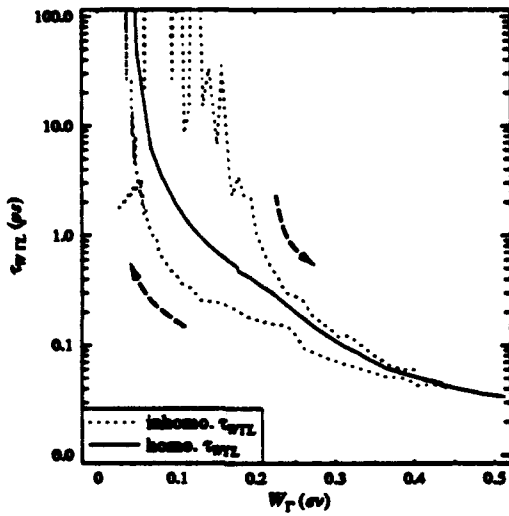


Figure 7: Energy relaxation time, τ_{WTL} , versus the average Γ -valley energy, W_Γ .

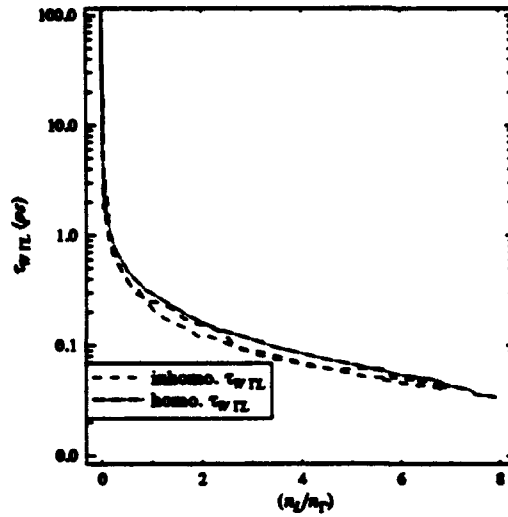


Figure 8: Energy relaxation time, τ_{WTL} , versus $(\frac{n_L}{n_T})$.

Computation of the Three-dimensional Depletion Approximation by Newton's Method and Multigrid

Franklin Bodine, Michael Holst and Thomas Kerkhoven

Beckman Institute and Department of Computer Science
University of Illinois at Urbana-Champaign
Urbana, IL 61801,
USA

Abstract

The three-dimensional (3D) depletion approximation is computed numerically by solution of a modified version of Poisson's equation for the electrostatic potential. Nonlinearities are resolved by a globally convergent modification of Newton's method. The linear systems are solved by 3D multigrid routines.

I. Introduction

Computational algorithms to resolve the nonlinearities in numerical drift-diffusion modeling such as Newton's or Gummel's method generally require a reasonably accurate initial guess. The charge neutral approximation provides one possible option that is obtained with little computational effort. An alternative approximate solution is provided by the depletion approximation, which tends to be presented in the context of analytical one-dimensional modeling. In this paper we present numerical results obtained with a novel computational multi-dimensional implementation of the depletion approximation.

If q is the size of the electron charge, T the ambient temperature, and k_B Boltzmann's constant, then the thermal potential $U_T \equiv (k_B T)/q$. The intrinsic Debye length $\lambda_{D,i}$ is then equal to $\lambda_{D,i} \equiv \sqrt{(\epsilon_r \epsilon_0 k_B T)/(q^2 n_i)}$. We express length in terms of an arbitrary unit l_0 . The nonlinear Poisson equation, in terms of the electrostatic potential u (in units of the thermal U_T) and the quasi-Fermi levels v and w , is then given by

$$-\nabla^2 u + \left(\frac{l_0}{\lambda_{D,i}}\right)^2 [e^{u-v} - e^{w-u} - k_1] = 0. \quad (1)$$

In thermal equilibrium the quasi-Fermi levels v and w may be set equal to the constant value 0. In the charge-neutral approximation the potential function u is set equal to

$$u_{cn} \equiv \sinh^{-1}(k_1/2). \quad (2)$$

However, it is obvious that the charge neutral approximation u_{cn} is significantly less accurate as an estimate to the solution u to Eq. (1) if v and w vary significantly, reflecting applied bias potentials. The charge-neutral approximation will be particularly inaccurate in modeling the depletion around reverse biased p-n junctions.

Our algorithm to compute the multi-dimensional depletion approximation employs a modification of the nonlinear Poisson equation (1) in which we attempt to reduce mobile charge

densities to 0 where the solution is furthest from thermodynamic equilibrium. The modified Poisson equation is given by

$$-\nabla^2 u_{depl} + \left(\frac{l_0}{\lambda_{D,i}}\right)^2 [(e^{u_{depl}-v_{pc}} - e^{w_{pc}-u_{depl}}) + e^{-\sqrt{(u_{depl}-v_{pc}-(u_{te}-v_{te}))^2+1}} + 1 - k_1] = 0 \quad (3)$$

where u_{depl} is the potential distribution which accounts for all depletion, v_{pc} and w_{pc} are the piece-wise constant quasi-Fermi levels, u_{te} is the thermal equilibrium potential distribution, and v_{te} is the quasi-Fermi level in thermal equilibrium (which is zero). The nonlinearities are resolved with Newton's method and the linear systems are solved with three-dimensional multigrid routines developed by the second author [1]. Special care was taken to keep the linear systems well-conditioned. First, we solve Eq. (1) without externally applied bias for u_{te} . Then the externally applied biases are applied at the contacts. The quasi-Fermi levels are adjusted in a piece-wise constant manner in each doping region to match the applied bias for the contact on that region. Then Eq. (3) is solved for the complete depletion.

II. Numerical Software

The linear multilevel method at the core of the software, as described in [1, 2], allows for discontinuous coefficients as occur in material interface problems. Operator-induced prolongation procedures are used to enforce flux conservation at box boundaries when a mesh function is interpolated from a coarse to a fine mesh. Smoothing operators are Red/Black Gauss-Seidel and weighted Jacobi, whereas the coarse mesh problem is solved with conjugate gradient methods. Nonlinear problems are handled with an extremely robust globally convergent damped-inexact-Newton-multilevel solver based on fast linear multilevel methods for the inexact Jacobian system solves. A nonlinear prolongation operator has been employed for nested Newton iteration. The discretization is with the box method.

To handle the severe numerical problems occurring with nonlinearities of exponential-type present in the semiconductor equations, we developed argument-capping functions which avoid nonvectorizable statements. Calls to the standard intrinsic functions are replaced by these modified functions, and overflows are successfully avoided during early transient iterations without losing the execution efficiency of the intrinsic functions. Due to the various choices made during the development of this package, the software executes at very high rates on a number of modern computers; see for example [1, 2] for benchmarks.

III. Numerical Results

In the following pages, we show the initial charge neutral, Fig. 1, the thermal equilibrium, Fig. 2, and the final depleted, Fig. 3, electron depletion for a three-dimensional BJT model. The computational grid was $49 \times 49 \times 25$ yielding a total of 60,025 unknowns. The computation of the our depletion model took 48 seconds on a HP 735 workstation, demonstrating the considerable computational efficiency of the multigrid routines.

Figures 1, 2, and 3 show two cut planes, one on the top on the of the device and one down the middle. The emitter is near the center, but slightly to the left, the base contact

Table 1: Timing and Iteration Statistics

Grid Size	Number of Points	Solution Time	Newton Iterations Thermal Equil.	Newton Iterations Applied Bias
25 × 25 × 25	15,625	15.02	5	6
25 × 25 × 33	20,625	20.04	5	6
33 × 33 × 33	35,937	40.16	5	7
49 × 49 × 33	79,233	103.07	6	7
49 × 49 × 49	117,649	152.23	6	7

is to the right of the emitter, and the collector contact is near the right edge. The lighter areas of the planes, representing lower electron densities ($\leq 5 \times 10^{16} \text{ cm}^{-3}$), are the parts of the base and collector which have been depleted of electrons.

The device simulated is a BJT built in an area of $1 \mu\text{m} \times 1 \mu\text{m}$ which is $0.5 \mu\text{m}$ deep. The device has a realistic doping profile [3, 4]. The background n-doping of 10^{17} cm^{-3} forms the collector. The base has a $0.46 \mu\text{m} \times 0.50 \mu\text{m}$ surface area and a depth of $0.15 \mu\text{m}$ with a p-type doping of 10^{18} cm^{-3} . The n-type emitter, which is doped at 10^{20} cm^{-3} , has a $0.08 \mu\text{m} \times 0.13 \mu\text{m}$ area and a depth of $0.04 \mu\text{m}$. This base width corresponds closely to [5].

The Debye length is given by [6], $L_D = \sqrt{(\epsilon_r \epsilon_0 k_B T) / (q^2 N_B)}$, where N_B is the doping concentration on the more lightly doped side of an abrupt junction. The Debye length of the emitter-base junction in this device is 4.1 nm and 13 nm for the collector-base junction.

The depletion width of a junction is given by [6] $W = L_D \sqrt{2[q(V_{bi} - V_{apl}) / (k_B T) - 2]}$, where V_{bi} is the built-in voltage and V_{apl} is the applied bias. The emitter-base junction has a depletion width of $0.036 \mu\text{m}$. The collector-base junction is $0.10 \mu\text{m}$ when no bias is applied and $0.22 \mu\text{m}$ when $V_E = V_B = 0 \text{ V}$ and $V_C = 3.0 \text{ V}$.

The metallurgical junction width from the bottom of the emitter to the bottom of the base region is $0.103 \mu\text{m}$. The calculated electron depletion region (the base width plus the collector-base depletion, which is assumed to take place entirely in the collector) is $0.208 \mu\text{m}$ for the zero bias case while it is $0.327 \mu\text{m}$ for the biased case. Using a threshold of a factor of two less than the doped concentration as the edge of the depletion region, the results from the thermal equilibrium solution gave an electron depletion region of $0.195 \mu\text{m}$ while the depletion calculation gave $0.303 \mu\text{m}$. These compare favorably to the analytically calculated results; however, agreement will not be exact due to three dimensional effects and the error in assuming all of the depletion occurs on the more lightly doped side of the junction.

The width of the undepleted base decreased from $0.103 \mu\text{m}$ in the charge neutral case to $0.072 \mu\text{m}$ at thermal equilibrium. When the final depletion correction including the bias was added, the base width was $0.059 \mu\text{m}$. These results are similar to experimental measurements [7].

The comparison of the charge-neutral approximation u_{cn} and the multi-dimensional depletion approximation u_{depl} demonstrates clearly the superiority of the computational depletion approximation. The strong rounding of the electron depletion region in our device with an exactly rectangular doping profile, demonstrates the need for a 3-D calculation. Table 1 reflects the linear computational complexity of the calculation.

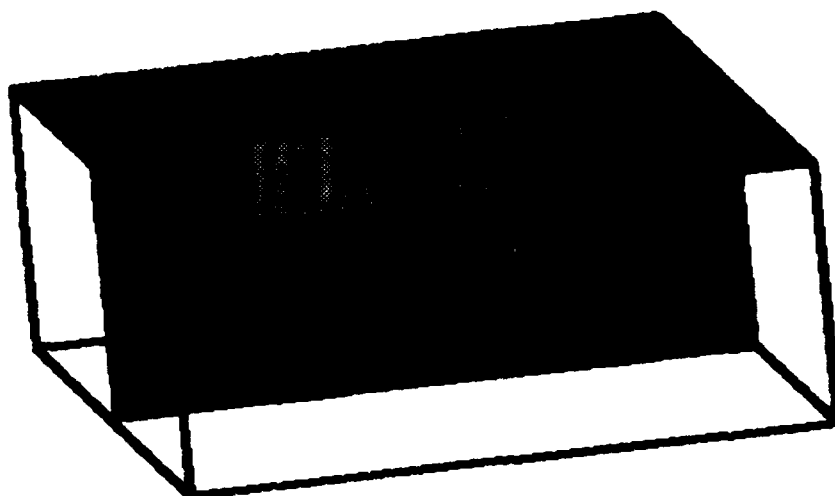


Figure 1: Charge neutral solution. The light area of both planes is the base region while the exterior dark region is the collector and the interior dark region is the emitter. This is the same orientation as the above figure: looking up from slightly below the device. The contacts are on the top plane.

References

- [1] Michael Holst. *Multilevel Methods for the Poisson-Boltzmann Equation*. PhD thesis, Numerical Computing Group, Department of Computer Science, University of Illinois at Urbana-Champaign, 1993.
- [2] Michael Holst and Faisal Saied. Multigrid solution of the Poisson-Boltzmann equation. *Journal of Computational Chemistry*, 14(1):105–113, 1993.
- [3] Richard S. Muller and Theodore I. Kamins. *Device Electronics for Integrated Circuits*. John Wiley and Sons, New York, second edition, 1986.
- [4] S. M. Sze, editor. *High-speed Semiconductor Devices*. Wiley-Interscience, New York, 1990.
- [5] D. D. Tang and Paul M. Solomon. Bipolar Transistor Design for Optimized Power-Delay Logic Circuits. *IEEE Journal of Solid-State Circuits*, SC-14(4):679–684, Aug 1979.
- [6] S.M. Sze. *Physics of Semiconductor Devices*. Wiley-Interscience, New York, second edition, 1981.
- [7] Joachim N. Burghartz, Jack Yuan-Chen Sun, Carol L. Stanis, Siegfried R. Mader, and James D. Warnock. Identification of Perimeter Depletion and Emitter Plug Effects in Deep-Submicrometer, Shallow-Junction Polysilicon Emitter Bipolar Transistors. *IEEE Transactions on Electron Devices*, 39(6):1477–1489, Jun 1992.

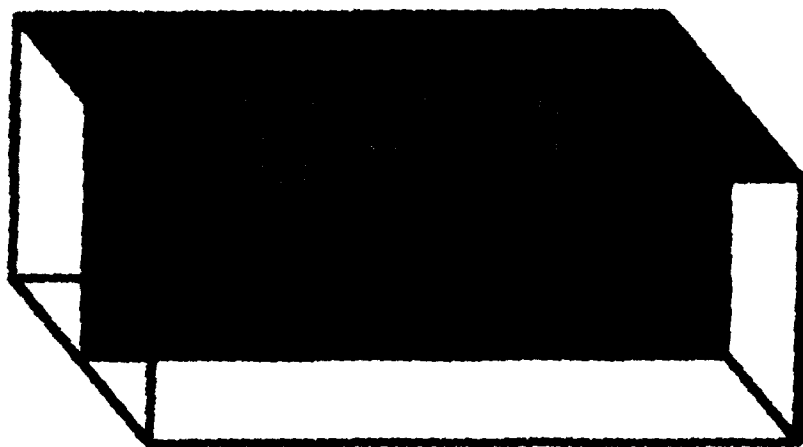


Figure 2: Thermal equilibrium solution. The lighter area is the region of the base which is not depleted in the thermal equilibrium case. The device is $1\ \mu\text{m}$ on a side.

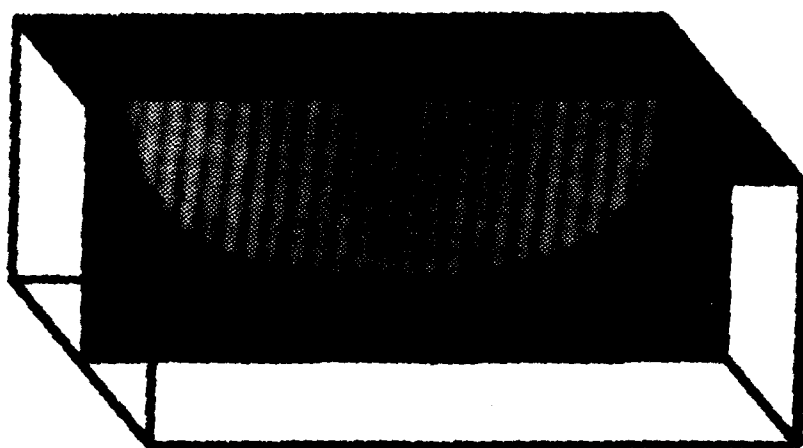


Figure 3: Depletion approximation solution. The small light area is the undepleted region of the base when the contact potential is applied.

An Accurate and Efficient Method to Characterize Heterostructure Devices

Ali Abou-Elnour and Klaus Schuenemann

Technische Universitaet Hamburg-Harburg
Postfach 90 10 52, D-2100 Hamburg 90, Germany

Abstract

Electronic states inside heterostructure devices are obtained by solving Poisson's and Schroedinger's equation self-consistently. Schroedinger's equation is efficiently solved by using variational techniques to obtain the wave functions in terms of a number of expansion functions. The present method is used to characterize the operation of single-well and pseudomorphic heterostructures, to characterize quantum wire devices, to calculate the two-dimensional scattering rates, and in Monte-Carlo codes.

I Introduction

The physical phenomena which are taking place in ultra-small and highly doped heterostructure devices require accurate physical models to characterize device operation and to optimize the structures. The main difficulty is that the behavior in the two dimensional electron gas (2DEG) region depends in a complicated manner on the device geometry and on bias conditions. The self-consistent solution of Poisson's and Schroedinger's equations is one of the most accurate models which is used to characterize heterostructure semiconductors to overcome these difficulties [1,2].

Previous self-consistent calculations generally used the finite difference technique to solve Schroedinger's equation [3-5]. The mesh size and the discretization method deteriorated the accuracy of the obtained results. We have recently introduced an efficient method to solve Schroedinger's equation by using variational techniques [6]. In the present work, we apply this method to characterize different heterostructure devices.

II Modeling

The effective mass, one-dimensional Schroedinger equation is given by

$$-\left(\hbar^2 / 2 m^*\right) \left(\partial^2 \psi_i(x) / \partial x^2\right) + V(x) \psi_i(x) = E_i \psi_i(x) \quad (1)$$

where $V(x)$ means potential energy, E_i eigenenergy, $\psi_i(x)$ wave function corresponding to the eigenenergy E_i , m^* effective mass, and \hbar Planck's constant. For a semiconductor structure of width a , the wave functions can be expanded as

$$\psi_k = \sum_{n=1}^N a_{nk} \sin\left(\frac{n\pi x}{a}\right) \quad (2)$$

If N is infinite, the obtained wave functions are identical with the true ones. However, a finite N still leads to very good accuracy. The coefficients a_{nk} , the eigenenergies, and the corresponding wave functions are determined by solving the matrix equation resulting from the variational integral for E_k [6].

III Application

The previous method is applied to characterize a modulation doped structures by solving Poisson's and Schroedinger's equations self-consistently. The 2DEG density, the energy levels, the wave functions, and the transconductance are displayed versus the total gate voltage in figs. 1,2,3, and 4 respectively. The total gate voltage represents the sum of the applied external voltage and the Schottky barrier equivalent voltage. Good qualitative agreement between the calculated results and those from both experiments and Monte-Carlo calculations [7] is obtained.

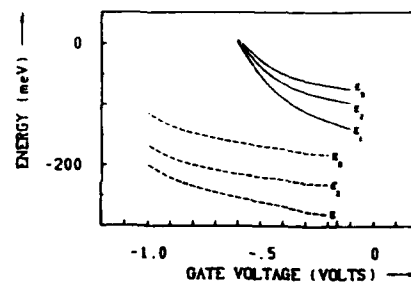
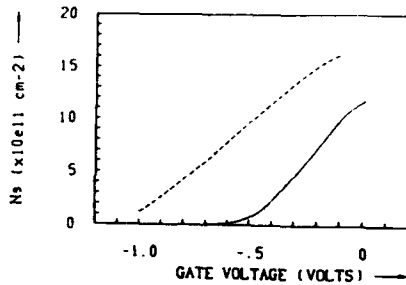
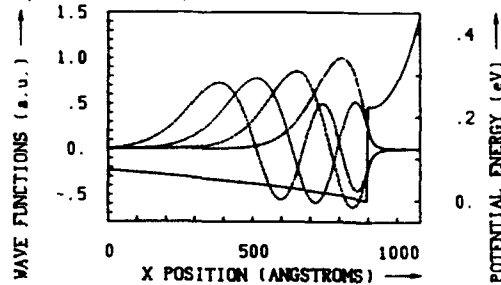
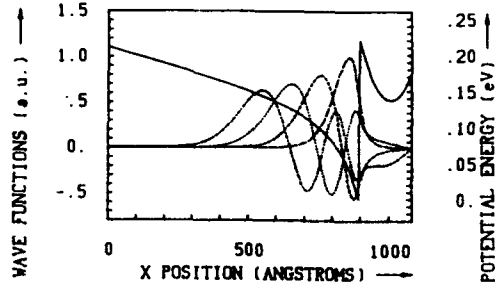
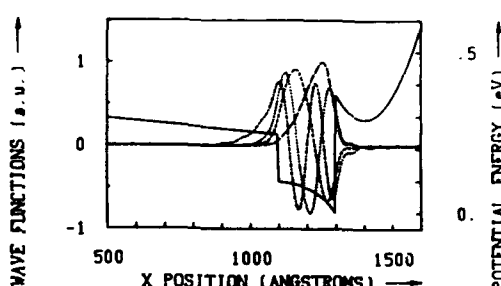
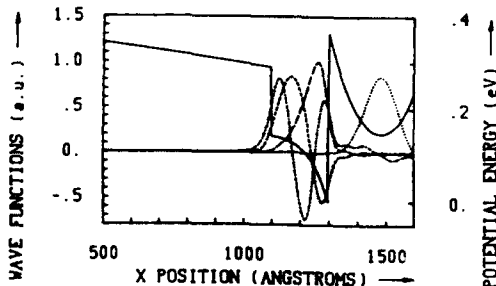


Fig.1 2DEG density versus gate voltage in single-well (solid line) and pseudomorphic (dashed line) structures. Fig.2 Subband energies versus gate voltage in single-well (solid lines) and pseudomorphic (dashed lines) structures.



(a) Single-well ($V_g = -0.1$ Volt).

(b) Single-well ($V_g = -0.5$ Volt).



(c) Pseudomorphic ($V_g = -0.1$ Volt).

(d) Pseudomorphic ($V_g = -0.5$ Volt).

Fig.3 Energy-band (dashed line) and wave functions (solid lines) for

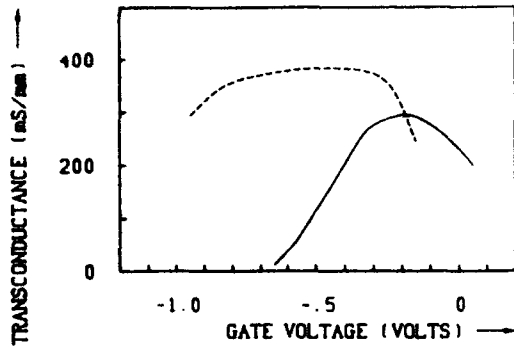


Fig.4 Transconductance versus gate voltage in single-well (solid line) and pseudomorphic (dashed line) structures.

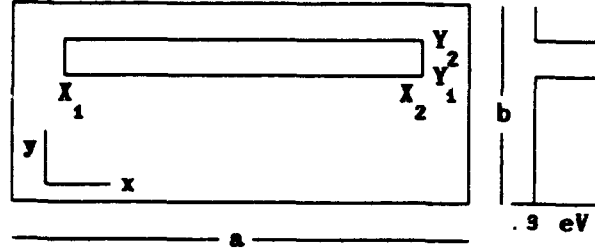


Fig.5 Schematics of the simulated structure. $a=b=100\text{nm}$, $X1=20\text{nm}$, $X2=80\text{nm}$, $Y1=60\text{nm}$, $Y2=70\text{nm}$

Quantum wire structures, shown in fig.5, can also be characterized by solving Schroedinger's equation in two dimensions using Rayleigh-Ritz method to obtain the two-dimensional wave functions (fig.6). The present method has the advantage that it requires reasonable CPU time, it is straight forward, and it overcomes the limitations of previous models arising from mesh size and discretization.

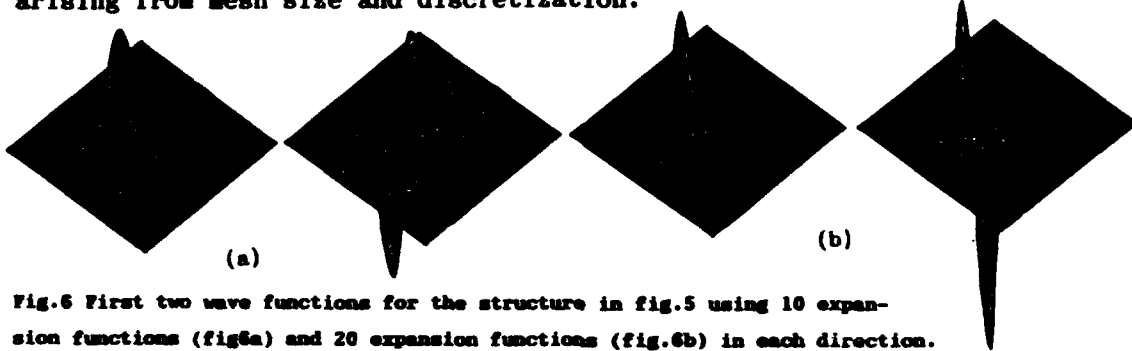


Fig.6 First two wave functions for the structure in fig.5 using 10 expansion functions (fig.6a) and 20 expansion functions (fig.6b) in each direction.

IV Calculation of the two-dimensional scattering rates

The two-dimensional scattering rates are calculated by defining the matrix element for scattering between the i th and the j th subbands according to

$$|M_{ij}|^2 = \int |M(Q,q)|^2 |I_{ij}(q)|^2 dq \quad (3)$$

where Q, q are the phonon wave-vector components parallel and normal to the hetero-interface, and $I_{ij}(q)$ means overlap integral

$$I_{ij}(q) = \int \psi_i(x) \psi_j(x) \exp(iqx) dx. \quad (4)$$

$\psi(x)$ is the normalized envelope wavefunction. Both Rayleigh-Ritz and finite difference methods are applied to calculate the subband energies and the corresponding wavefunctions of an AlGaAs/GaAs heterostructure. Using Rayleigh-Ritz method, the required CPU time (fig.7b) to calculate the 2D scattering rates (fig.8) versus the number of subbands is nearly constant while it greatly changes using the finite difference method (fig.7a). This makes the application of the present method more practical in particular for device simulation.

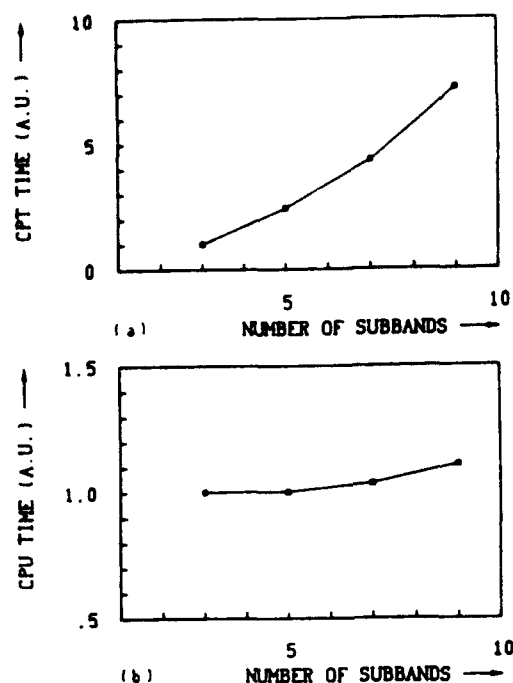


Fig. 7 Required CPU time to calculate the 2D scattering rates versus number of subbands. a) finite difference, b) Rayleigh-Ritz method.

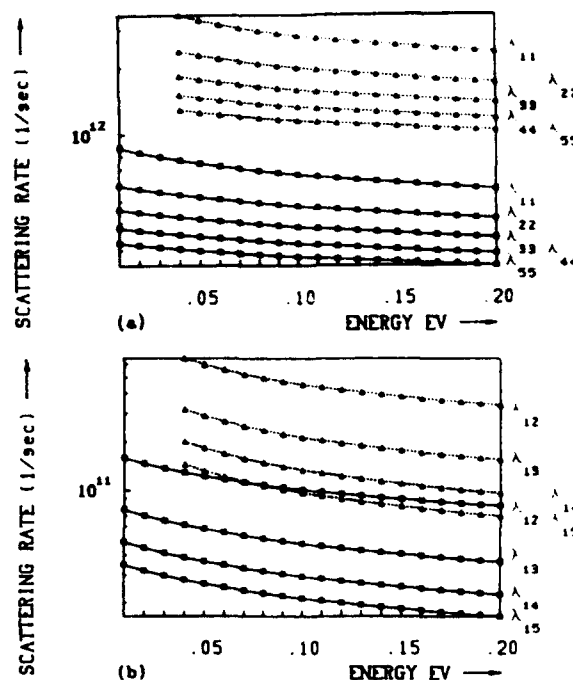


Fig. 8 2D polar optical phonon scattering rate absorption (solid lines), emission (dashed line) a) intrasubband, b) intersubband scattering.

V Monte-Carlo simulation of Hetero-FETs

A two-dimensional Monte-Carlo code is investigated to simulate the hetero-FET structure shown in fig. 9. Poisson's and Schroedinger's equations are solved self consistently along the device and the obtained results are used to express the conduction band by a number of step functions. The wave functions and equivalent energy wells at different sections along the channel are displayed in fig. 10.

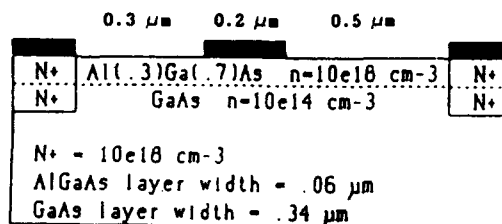


Fig. 9 Simulated AlGaAs/GaAs Hetero-FET.

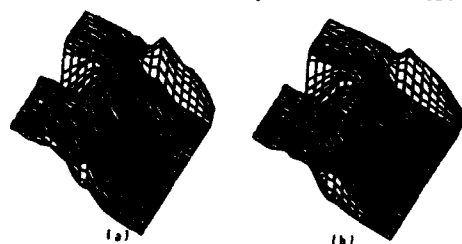


Fig. 11 Carrier distribution in the channel (a) the present model (b) classical model.

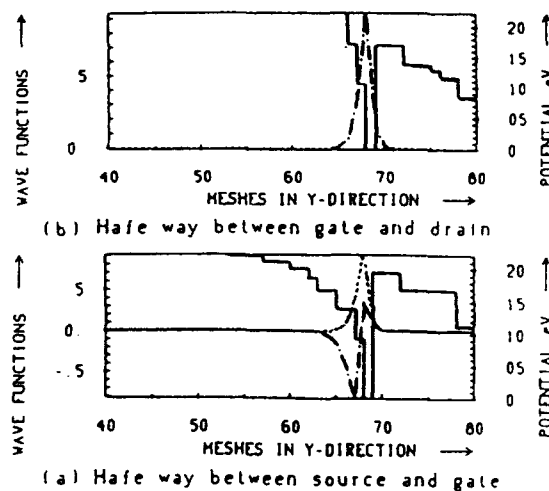


Fig. 10 Energy wells (solid lines) and wave functions (dashed lines) at different positions [mesh size 50A].

The present model takes the size quantisation into account because the electrons which have energies less than the barrier height can only move in two dimensions. Moreover, the transfer between the different regions is simpler than in other models because no extra scattering rates are required to be derived. Our model leads to a higher carrier concentration in the channel region than the classical models, fig. 11, because the carrier capture in the 2DEG region is better simulated.

VI Conclusions

An efficient computational tool is used to determine the electronic states inside heterostructure. The present model overcomes the limitations of the previous models which arise from mesh size and discretization. The closed forms of the wave functions makes the calculations of the two-dimensional scattering rates easier. This method can also be efficiently used together with Monte-Carlo codes to characterize semiconductor devices.

Acknowledgement

The authors are thankful to Prof. Dr. A. S. Omar for fruitful discussions and to the Deutsche Forschungsgemeinschaft for financial support.

References

- [1] K. Hong and D. Pavlidis, "Self-consistent analysis of lattice-matched and pseudomorphic quantum-well emission transistor", J. Appl. Phys. 69, 2262 (1991).
- [2] K. Yokoyama and K. Hess, "Monte-Carlo study of electronic transport in $\text{Al}_{(1-x)}\text{Ga}_x\text{As}/\text{GaAs}$ single-well heterostructure" Phys. Rev. B 33, 5595 (1986).
- [3] M. Patil and H. Morkoc, "Self-consistent calculation of electron density in a two-channel modulation doped structure", Solid-State Elec. 33, 99 (1990)
- [4] I-H. Tan, G. Snider, L. Chang, and E. Hu, "A self-consistent solution of Schrodinger-Poisson equations using a nonuniform mesh" J. Appl. Phys. 68, 4071 (1990).
- [5] A. Curz Serra and A. Abreu Santos, "A one-dimensional, self-consistent numerical solution of Schrodinger and Poisson equations" J. Appl. Phys. 70, 2734 (1991).
- [6] A. Abou-Elnour and K. Schuenemann, "Accurate and efficient modeling of Hetero-FETs", Proc. IEEE MTT-Symposium, 1177 (1993).
- [7] D. Park and K. Brennan, "Monte Carlo simulation of $0.35\ \mu\text{m}$ gate-length GaAs and InGaAs HEMT's" IEEE Trans. Electron Devices ED-37, 618 (1990).

A New, Easy-to-Code, Robust and Stable Approach to 2-D Hydrodynamic Submicron Device Modeling

Qi Lin, Neil Goldsman and Gwo-Chung Tai
Department of Electrical Engineering
University of Maryland
College Park, MD, 20742

I. Introduction

We present a new technique for solving the Hydrodynamic(HD) equations in submicron device simulations. This method is extremely stable, quickly converges even with poor initial guesses, and agrees with Monte Carlo(MC) calculations. In addition, this new method is easy to code for 2-D device simulation. We have applied the new method to simulate 2-D MOSFETs as well as SOI devices.

To our knowledge, no agreed-upon HD model has emerged as the basis for a standard device simulator. The lack of an industry-standard CAD tool can be attributed to the difficulties in obtaining an accurate, numerically stable, and rapidly convergent solution to the HD equations. We have developed a new robust algorithm for HD device simulation that overcomes existing difficulties. We formulate the HD equations into self-adjoint forms with a new set of Slotboom-like state variables. The discretizations result in a diagonally dominant coefficient matrix for each HD equation. Consequently, the convergence of each equation is guaranteed for any initial guess when iterative solution methods are employed. Our discretization technique resolves the rapid spatial variations which may occur in carrier densities and carrier temperatures. As a result, stability of the HD solution is improved. Furthermore, a fixed-point iterative method is employed to determine the solution of each discretized HD equation. A direct solution of a matrix equation is therefore avoided. In addition, the method requires little memory, and is well-suited for parallel computations.

II. The HD Model

The HD equations are obtained from a standard HD formulation [1].

$$\nabla^2 \phi = \frac{q}{\epsilon_s} (n - p - D) \quad (1)$$

$$\frac{1}{q} \nabla \cdot \vec{J}_n = R \quad (2)$$

$$\nabla \cdot \vec{S}_n = \vec{J}_n \cdot \vec{E} - n \frac{\omega_n - \omega_o}{\tau_{nw}(T_n)} \quad (3)$$

$$\vec{J}_n = -q\mu_n n \nabla \phi + qD_n \nabla n + \mu_n k_B n \nabla T_n \quad (4)$$

$$\vec{S}_n = \frac{\vec{J}_n}{-q} \omega_n + \frac{\vec{J}_n}{-q} k_B T_n + \vec{Q}_n \quad (5)$$

$$\omega_n = \frac{2}{3}k_B T_n + \frac{1}{2}m_n^* v_{dn}^2 \quad (6)$$

$$\vec{Q}_n = -\kappa_{cn} \nabla T, \quad \kappa_{cn} = 2T_n(k_B/q)^2(nq\mu_n) \quad (7)$$

Here, ϕ is the electric potential; n and p are the electron and hole densities respectively; D is the net doping concentration; \vec{J}_n is the electron current density; R denotes the recombination process; ω_n is the average electron energy and \vec{S}_n is the electron energy flux; \vec{E} is the electric field; τ_{nw} is the energy relaxation time for electrons; ω_o is the average carrier energy in thermal equilibrium; μ_n is the electron-temperature-dependent mobility; D_n is electron temperature-dependent diffusivity; T_n is the electron temperature; m_n^* is electron effective mass; v_{dn} is the electron mean velocity; \vec{Q} is the heat flux. (We note that, in an effort to be concise, only the HD equations for electrons were shown. However, our calculations include the self-consistent solution of the hole and electron HD equations.)

The typical approach at this point is to solve the above system of equations by using n, T_n as the unknown state variables. However, the current-continuity and the energy-balance equations under this approach may give rise to numerical difficulties such as stability problems and spurious spikes in average electron velocity. To overcome these numerical problems, we take another approach by first defining a new set of Slotboom-like state variables. Then, we transform the HD equations into self-adjoint forms with these new variables. A new Scharfetter-Gummel-like discretization scheme is then employed to resolve the rapid variations in n and T_n . The resulting matrix equations are diagonally dominant and exhibit excellent numerical properties.

III. The New Method for Solving HD Equations

A. Slotboom-Like Variables for the HD Model

We define a new set of Slotboom-like state variables u and g_n for electron density and electron temperature:

$$n = n_i \exp\left(-\frac{T_n - T_L}{T_L} + \frac{q(\phi - \psi_n)}{k_B T_L}\right) = n_i u \exp\left(-\frac{T_n - T_L}{T_L} + \frac{q\phi}{k_B T_L}\right) \quad (8)$$

$$u = \exp\left(-\frac{q\psi_n}{k_B T_L}\right) \quad (9)$$

$$T_n = g_n \exp(\psi_n/a_T), \quad a_T = \frac{\kappa_{cn}}{\frac{5}{2}\mu_n n k_B} = \frac{4}{5} \frac{k_B T_L}{q} = 0.0207V \quad (10)$$

where ψ_n is the electron quasi-Fermi potential and T_L is the lattice temperature.

B. The Self-Adjoint form of the HD model

Substituting eqns. (8)(9) into (4), and eqn. (10) into (5) respectively, the electron current density and electron energy flux in terms of the Slotboom-like variables u and g_n are as follows:

$$\vec{J}_n = D_n n_i \exp\left(-\frac{T_n - T_L}{T_L} + \frac{q\phi}{k_B T_L}\right) \nabla u, \quad (11)$$

$$\vec{S}_n = -\kappa_{cn} \exp(\psi_n/a_T) \nabla g_n + \frac{\vec{J}_n}{-q} \left(\frac{1}{2} m_n^* v_{dn}^2\right). \quad (12)$$

Eqn. (11) can be further simplified to $\vec{J}_n = -q\mu_n n \nabla \psi_n$. The steady state HD model can now be described by a system of self-adjoint equations. This system can be obtained by appropriately substituting eqns. (11)-(12) into eqns. (1)-(3).

$$\nabla^2 \phi = \frac{qn_i}{\epsilon_s} \left(u \exp\left(-\frac{T_n - T_L}{T_L} + \frac{q\phi}{k_B T_L}\right) - v \exp\left(-\frac{T_p - T_L}{T_L} - \frac{q\phi}{k_B T_L}\right) \right) - \frac{qD}{\epsilon_s} \quad (13)$$

$$\nabla \cdot \left(D_n n_i \exp\left(-\frac{T_n - T_L}{T_L} + \frac{q\phi}{k_B T_L}\right) \nabla u \right) = R(\phi, u, v) \quad (14)$$

$$\nabla \cdot (\kappa_{cn} \exp(\psi_n/a_T) \nabla g_n) = \frac{n \frac{3}{2} k_B g_n \exp(\psi_n/a_T) + \frac{1}{2} m_n^* v_{dn}^2 - \omega_o}{\tau_{nw}} - \vec{J}_n \cdot \vec{E} - \nabla \cdot \left(\frac{\vec{J}_n}{q} \frac{1}{2} m_n^* v_{dn}^2 \right) \quad (15)$$

It is clear from the above expressions that the Poisson, the current-continuity and the energy-balance equations are each self-adjoint differential equations with respect to the variables ϕ, u and g_n . It is also interesting to note that when T_n is equal to T_L , the above new expressions for the Poisson and current continuity equations reduce to the DD model[2].

C. Discretization Scheme and Iterative Method for HD Equations

A Scharfetter-Gummel-like method is employed to discretize the current-continuity equations and the energy-balance equations. The discretizations result in a diagonally dominant coefficient matrix for each HD equation. A fixed-point method is applied to solve the system of discretized HD equations. Due to the property of diagonal dominance, the convergence for the solution of each HD equation is guaranteed [2,3].

IV. Numerical Results

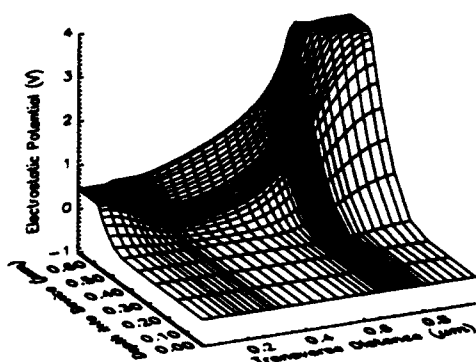
To test convergence of the technique, we generated initial guesses using a random function. With random initial guesses, the same results were obtained as when good initial guesses were used. To examine stability, we simulated MOSFET's on a rather coarse grid of less than 400 mesh-points and smooth results were obtained. Additionally, we found that, in general, our HD simulations required only about 75% more CPU time than drift-diffusion model simulations. Fig. 1 shows the simulation results for a submicron MOSFET with $0.5\mu m$ channel length. In Fig. 2, we show the results of a computation performed for a $1\mu m$ -channel SOI device. Finally, to test the accuracy of the new HD model, we compared our simulations to MC calculations for a square shape

field. Fig. 3 shows good agreement between MC simulations and our HD calculations. Also interesting in Fig. 3 is that the spurious second overshoot spike, which usually appears in HD simulations, is absent from our calculations.

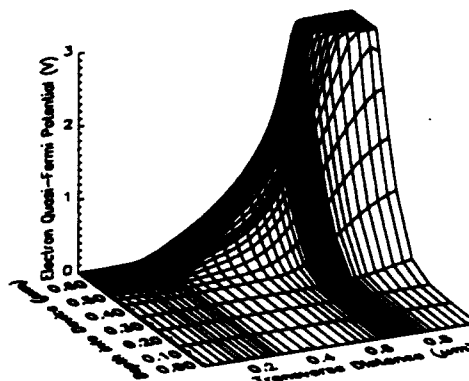
References

- [1] K. Blotekjaer, *IEEE Trans. Electron Dev.*, vol. ED-17, pp.38, 1970
- [2] C. Korman and I. D. Mayergoyz, *Journal of Appl. Phys.*, 68 (3), pp. 1324-1334, 1990
- [3] Q. Lin, N. Goldsman and G.-C. Tai, *Solid-State Electron*, vol. 36, no. 3, pp. 411-419, 1993.

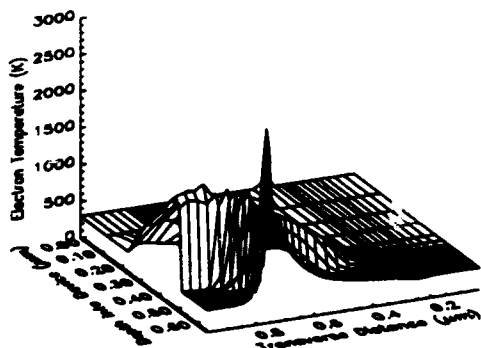
Figure 1. 2-D Submicron MOSFET Simulation Results
($V_{gs} = 3V$, $V_{ds} = 3V$, Channel Length = $0.5\mu m$)



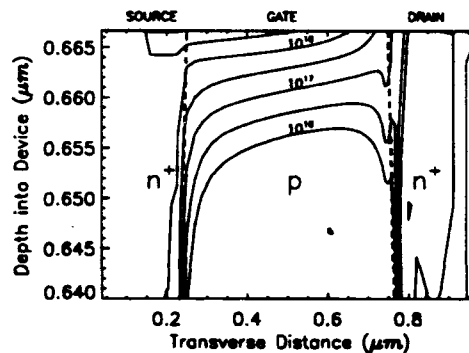
The Electrostatic Potential



The Electron Quasi-Fermi Potential



The Electron Temperature



The Electron Density

Figure 2. 2-D Partially Depleted SOI MOSFET Simulation Results
($V_{gs} = 2V$, $V_{ds} = 2V$, Channel Length = $1\mu m$)

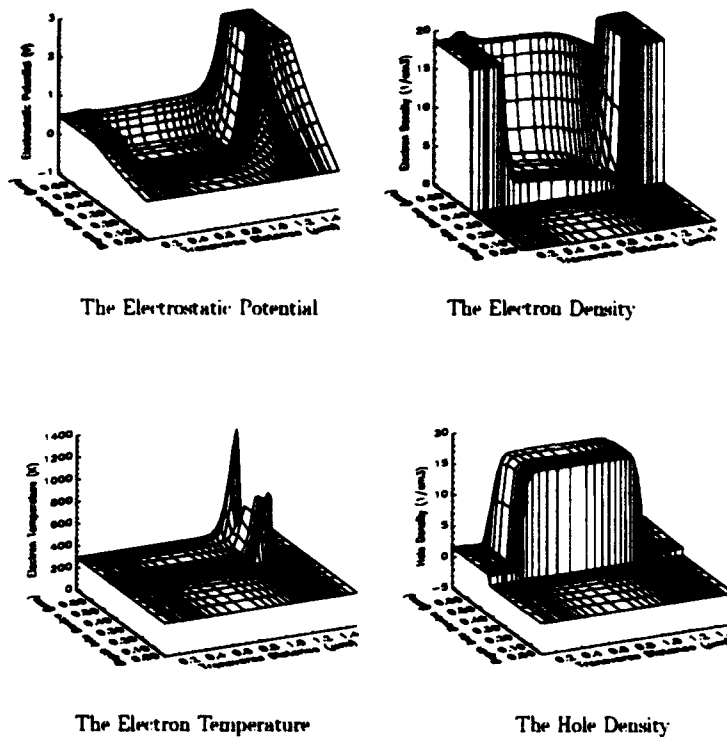
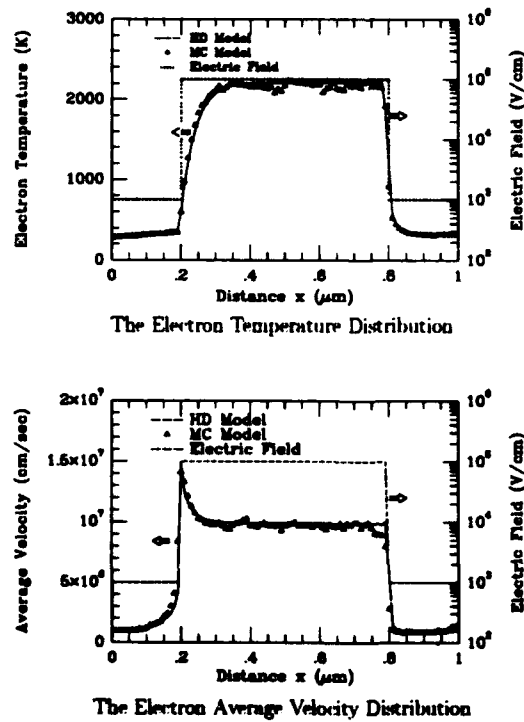


Figure 3. Comparison between HD and MC Calculations



An Equivalent Circuit Model and Distribution-Function Theory of the High-Frequency Behavior of Quantum-Based Devices

F. A. Buot and A. K. Rajagopal
Naval Research Laboratory
Washington, D.C. 20375

Abstract

The high-frequency response of resonant tunneling devices (RTD), subjected to time-dependent signal, is considered in both stable and unstable situations when operating in the negative differential resistance (NDR) region, based on the phase-space distribution function formalism. In the stable case, the equivalent-circuit approach (ECA) is shown to characterize the complex high-frequency behavior of a RTD's response to small a.c. signal. The ECA is found to be very useful in resolving the various outstanding controversies concerning the dynamical quantum transport behavior of RTD. For the unstable case, nonperturbative approaches are outlined.

I. Introduction

There is much confusion in the literature concerning the high-frequency behavior of resonant tunneling devices (RTD). The numerical simulation of Frensky [1], using the Wigner distribution-function transport equation (WDFTEQ), reveals a capacitive behavior at lower frequencies, eventually changing into an inductive behavior at higher frequencies. In contrast, the numerical simulation of Klusdahl, et al. [2] using a similar WDFTEQ approach reveals an inductive behavior at low frequencies, changing into a capacitive behavior at intermediate frequencies, and eventually changing back into a (somewhat) inductive behavior at high frequencies. Both of these simulations do agree concerning the real part of the admittance, namely, it is negative at lower frequencies and becomes positive at high frequencies. Their results also concur at high frequencies by having a cutoff in the admittance. Until now, the discrepancy in the reactive behavior cited above has remained unresolved.

So far, there are no other serious work on the subject of characterizing the high-frequency response of RTD to a.c. signal, although there have been a number of attempts to analyze the high-frequency behavior of nanometric structures. These are attempts to extend the Landauer-Büttiker viewpoint, which calculates conductance from transmission coefficients, to high frequencies. The gross deficiencies of all these attempts have been discussed by Landauer [3]. There are also attempts, for example by Fu and Dudley [4], which employ the machinery of the linear response theory, however this particular approach is plagued by its inability to treat far-from-equilibrium operating bias conditions, of interest to the nanoelectronics community. More recent results of Cai and Lax [5] employ a nonperturbative time-dependent Green's function approach to compute the behavior of an electron incident from an energy channel on a double-barrier structure. Their results reveal the presence of intrinsic oscillations in response to a voltage pulse across the negative differential resistance (NDR) region, similar to the result obtained by Buot and Jensen [6], particularly when self-consistency was switched off in their WDFTEQ calculations. The results of Cai and Lax are however limited to single-channel and nonselfconsistent contributions.

The purpose of this paper is twofold: to show that a coherent and unified understanding of the high-frequency behavior of RTD arises from the equivalent-circuit model of Buot and Jensen (BJ) [7], to be referred here as the BJ model (originally proposed to study the nonlinear aspects of RTD which arises from the selfconsistent treatment of WDFTEQ), and to point out a proper way to a nonperturbative theoretical framework for analyzing the high-frequency behavior of resonant tunneling devices within the quantum distribution-function approach. It is worthwhile to stress the first point here since there have been misleading statements in the literature claiming that no understanding exists for quantum transport at nonzero frequencies. Also there is a need to clarify a statement made in the literature which strongly implies that no LRC circuitry can simulate the complicated frequency dependence of the admittance in the linear-response regime, particularly since this statement is made without further reservations. We will here show that indeed the two earlier numerical simulations mentioned above [1,2] may be recast in terms of the BJ model in two separate regimes in the BJ circuit parametrization, operating in the NDR region. Thus, the BJ model is found to include the high-frequency behavior of RTD.

II. Equivalent-Circuit Model

The BJ model is derived from the selfconsistent quantum distribution-function (QDF) transport simulation, with four independent equivalent-circuit parameters, namely, series resistance, R , negative conductance, G , quantum inductance, L , and capacitance, C . The RTD simulated by BJ, with fixed bias applied in the NDR, is characterized by the following inequalities for the circuit parameters: $R|G| > 1$ and $RC > L|G|$. These are conditions for the presence of oscillatory behavior and instability of the operating point in the middle of the NDR region [7] leading to the intrinsic bistability and hysteresis in the current-voltage characteristics, which was the focus of the original BJ model. In contrast, we now find that the RTD simulated by Frensley [1] and Kluksdahl, et al. [2] are obtained if the following inequalities are obeyed: $R|G| < 1$, and $RC < L|G|$, along with further conditions to be specified in what follows. The above are conditions for the nonoscillatory and stable operation in the middle of the NDR region allowing for the applicability of linear-response approximation [8]. They also imply that the series resistance is quite small compared to that of the RTD simulated by BJ (note that the RTD simulated by BJ, although having approximately the same feature sizes for the double-barrier region, differs in having an undoped buffer layer on both sides of the double barrier, which may also affect the value of the capacitance relative to those of Refs. [1] and [2], and a longer computational box length, which together introduces a larger R). Moreover, the RTD simulated by Frensley [1], without taking self-consistency into account, differs from that simulated by Kluksdahl et al. [2], which takes self-consistency into account, by the following inequality: $L|G| < C/|G|$ for Frensley [1], whereas $L|G| > C/|G|$ for Kluksdahl et al [2]. These imply a larger quantum inductance for the RTD of Ref. [2] as compared to that simulated in Ref. [1], where the residual small inductance is mainly due to the quantum nonlocality [8]. Further, note that Ref. [2] uses wider barrier widths than those used in Refs. [1] and [7]. Thus the BJ circuit model serves not only to clarify the origin of the discrepancy of the results of the two numerical simulations [1,2], but also to bring together the various dynamical aspects of RTD in terms of WDFTEQ.

In addition to the quantum inductance, in parallel with the capacitance, found by BJ for fixed bias in the NDR region, it is important to take also into account effects arising from the electron kinetics in calculating the RTD response to high-frequency a.c. signal. These effects are often referred to as due to the electron inertia, i.e., it takes time for the electron to be accelerated and decelerated, typically causing the current to lag in time behind the electric field. These inertial

effects, although always present, are negligible at low frequencies. It is taken into account in our high-frequency equivalent circuit model of RTD by adding another inductance, ℓ , (ℓ is typically an order-of-magnitude smaller than the quantum inductance, L) in series with R , outside of the two-branch circuit of L , G , and C . As will be shown in the following figures, this additional "inertial inductance", ℓ , serves to cut off the RTD response at very high frequencies.

We have calculated the admittance of RTD to a small a.c. signal using the above-mentioned high-frequency equivalent circuit model. The different circuit parameters are estimated by adjusting the values obtained by BJ to conform with the constraints enumerated above appropriate to the different RTD simulated. The results are displayed in Fig. 1 for the RTD parameters appropriate to the device simulated by Frensky [1], and Fig. 2 for the RTD parameters appropriate to the device simulated by Klusdahl et al. [2] (note that the inductive part of the admittance is plotted with a positive scale in Ref. [2]). Observe that the characteristic features of the results of their numerical simulations, mentioned at the beginning, are well reproduced by the present results, obtained simply by using the high-frequency equivalent-circuit model. Indeed, these results lend further support to the accuracy of the BJ equivalent-circuit model of RTD at low frequencies [7], as well as confirm the validity of the present high-frequency equivalent circuit model of RTD which incorporates a series-inertial inductance. These results also serve to invalidate the claim [4] made without any reservation, that no LRC circuit can simulate the complicated frequency dependence of the admittance in the linear-response regime.

III. Nonperturbative Approaches for Unstable NDR Operation

For unstable operation in the NDR region, nonperturbative approaches are more appropriate. Here, a new time-dependent transformation of phase space is found which transform the quantum distribution (QDF) transport equation to the same form in the absence of time-dependent signal. This is shown to be a very powerful approach in revealing the nature of the nonperturbative response to a small a.c. signal. This time-dependent transformation is useful when the applied time-dependent electric field is assumed to be position-independent. The general formulation of the QDF transport equation in the presence of space and time-dependent potential is given by one of the authors [6,9]. For high-frequency signals applied at the drain terminal of an RTD, following the conventional procedure, we assume a time-dependent but position-independent applied perturbing field within the device. From the general formula given in Refs. [6,9], one arrives at a resulting QDF transport equation in response to this signal. We show that the resulting QDF transport equation can be transformed into the form of the equation obtained earlier in the absence of the perturbing time-dependent field, through a time-dependent transformation of the phase-space. This transformation of the equation however makes the double-barrier potential to be time-dependent. With this simplification, the application of the numerical technique of Ref. [6] with known time-dependent matrix for the double-barrier potential operator, with time as a parameter, becomes feasible. The derivation of the transformed and simplified new QDF transport equation, and the method of the numerical implementation for calculating the nonperturbative response to an a.c. signal will be discussed in a forthcoming paper.

In a more general and realistic situation, an applied time-dependent voltage at the drain will selfconsistently lead to a time-dependent and position-dependent potential inside the device. For this general situation of the unstable case, a nonperturbative approach based on the structure of phase space is proposed. We introduce two different representations of quantum transport, namely, the Liouville representation, and, the phase-space fluid representation. The QDF is

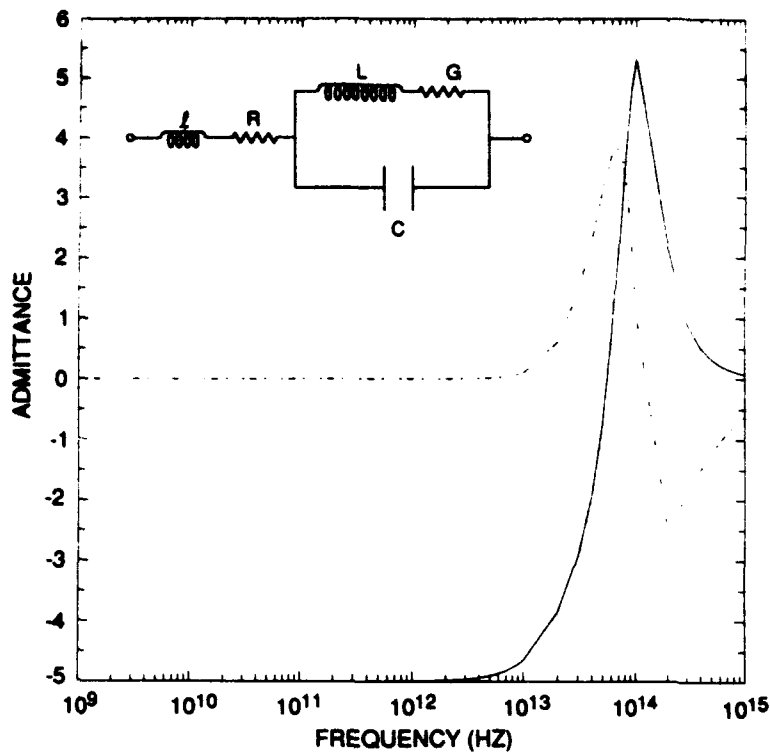


Fig. 1 - Real (solid curve) and imaginary (dashed curve) parts of the admittance (in units of 10^7 S/cm^2) for the circuit parameter combination that corresponds to the RTD simulated by Frenley [1]: $R|G| < 1$, $RC < L|G|$, and $L|G| < C|G|$. The values used are: $R = 0.2 \times 10^{-7} \Omega \text{ cm}^2$, $G = -2.5 \times 10^7 \text{ S/cm}^2$, $C = 6.0 \times 10^{-7} \text{ F/cm}^2$, $L = 0.8 \times 10^{-2} \text{ H/cm}^2$, and $l = 0.2 L$.

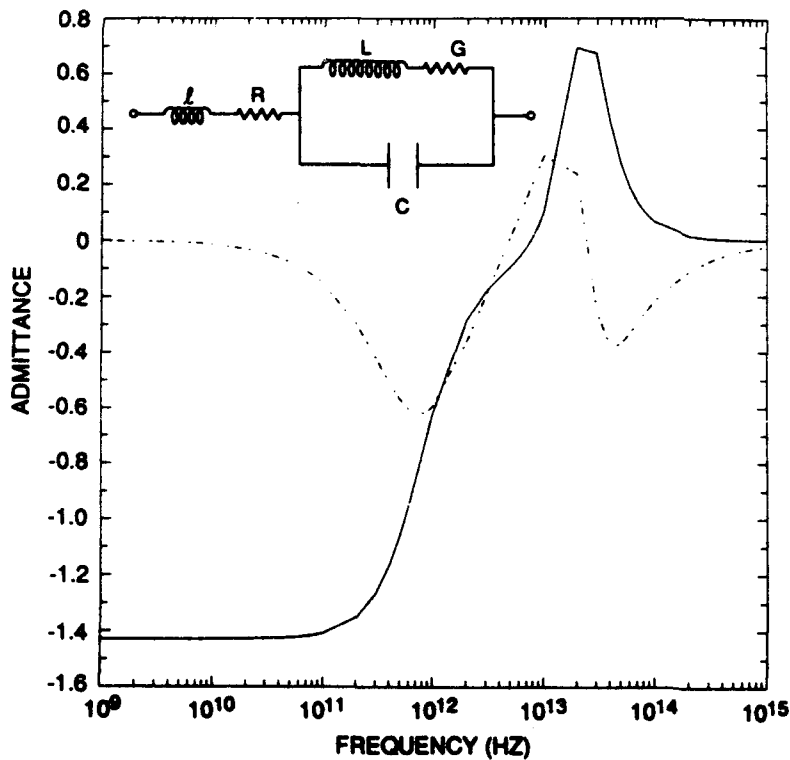


Fig. 2 - Real (solid curve) and imaginary (dashed curve) parts of the admittance (in units of 10^7 S/cm^2) for the circuit parameter combination that corresponds to the RTD simulated by Klukadahl et al. [2]: $R|G| < 1$, $RC < L|G|$, $L|G| > C|G|$. The values used are: $R = 1.3 \times 10^{-7} \Omega \text{ cm}^2$, $G = -0.5 \times 10^7 \text{ S/cm}^2$, $C = 0.4 \times 10^{-6} \text{ F/cm}^2$, $L = 9.0 \times 10^{-20} \text{ H/cm}^2$, and $l = 0.05 L$.

solved numerically in the same fashion as described and successfully implemented in the existing literature, but using time-dependent bias. The QDF solution has inherent undesirable features for studying the dynamics of phase space, which can be eliminated by a special post-processing. This post-processing yields a smoothed-out QDF, the positive definite Husimi distribution, and allow us to study the structure of quantum trajectories in phase space. It is suggested that the use of Husimi distribution enables a microscopic dynamical viewpoint of ECA as well as shed further light on the dynamical nature of the quantum inductance. The full details of this approach will be discussed in another paper.

IV. Concluding Remarks

In conclusion, the utility of the QDF approach for understanding the high-frequency behavior of quantum-based devices and in deducing the equivalent circuit model of RTD have been demonstrated here. It is worthwhile to reiterate that the equivalent circuit model of RTD is subtle and in the form presented here elucidates the two earlier simulations [1,2] as special cases of the model presented here. Thus the discrepancy in the simulated reactive behavior mentioned in the opening paragraph is resolved. Moreover, the exact role of electron inertia in the high-frequency behavior of RTD is clarified. For cases when the linear response approximation breaks down, nonperturbative schemes are outlined.

Acknowledgment

The partial support from the Office of Naval Research is gratefully acknowledged.

References

- [1] W.R. Frensley, Rev. Mod. Phys. **62**, 765 (1990). In this paper, the author corrects his earlier work (Appl. Phys. Lett. **51**, 448 (1987)) on the reactive behavior of RTD.
- [2] N.C. Kluksdahl, A.M. Kriman, D.K. Ferry, and C. Ringhofer, IEEE Elec. Dev. Letts. **9**, 457 (1988).
- [3] R. Landauer, Physica Scripta T**42**, 112 (1992).
- [4] Y. Fu and C. Dudley, Phys. Rev. Letts. **70**, 65 (1993).
- [5] W. Cai and M. Lax, Phys. Rev. **B47**, 4096 (1993).
- [6] F.A. Buot and K.L. Jensen, Phys. Rev. **B42**, 9429 (1990).
- [7] F.A. Buot and K.L. Jensen, Intl. J. Comp. Math. Electron. Electrical Eng., COMPEL **10**, 241 (1991).
- [8] F.A. Buot, "*Mesoscopic Physics and Nanoelectronics: Nanoscience and Nanotechnology*", to appear in Physics Reports (1993).
- [9] F.A. Buot, J. Stat. Phys. **61**, 1223 (1990).

Device simulation by means of a direct solution of the coupled Poisson/Boltzmann Transport equations

Conor J. Donnelly and Colin Lyden

National Microelectronics Research Centre,
University College, Cork, Ireland

Abstract

A method for directly solving the coupled Poisson/Boltzmann Transport equation including a new model for the collision term is presented. This method yields the complete momentum distribution function from which all quantities of interest to the device modeller can easily be calculated. Current continuity is implicit in the model. The method is demonstrated by simulating an n^+-n-n^+ silicon structure.

I Introduction

With decreasing device dimensions there is a need to accurately model the effects of hot carriers. Solutions of the drift-diffusion and hydrodynamic models have been widely used to model the behaviour of many submicron devices [1]. However, in the analysis of the programming of EPROMs [2] and many other phenomena found in modern small geometry devices, the complete carrier distribution function in both ordinary space and momentum space is required for accurate simulation. Recently there has been increased interest in solving the Boltzmann Transport Equation (BTE) to obtain the distribution function [3, 4]. Knowing the distribution function and electric potential, all other quantities of interest can be calculated. In this paper a method for solving the coupled Poisson/Boltzmann Transport equations is presented.

II Mathematical model

The BTE for steady state conditions for a semiconductor device is

$$\vec{v} \cdot \vec{\nabla}_r f - \frac{q}{\hbar} \vec{E} \cdot \vec{\nabla}_k f = - \left(\frac{\partial f}{\partial t} \right)_{COLL} \quad (1)$$

where \vec{v} is the carrier velocity, \vec{E} is the electric field and f is the distribution function. The right hand side of equation 1 is the collision term and is often approximated using the relaxation time approximation [5] as

$$\left(\frac{\partial f}{\partial t} \right)_{COLL} \simeq \frac{f - f_0}{\tau} \quad (2)$$

where f_0 is the distribution function at equilibrium and τ is the relaxation time. Integrating equation 1 over momentum gives the current continuity equation in the absence of generation-recombination. However, if the relaxation time approximation is used the integral of the collision term over momentum is, in general, not equal to zero because the electron concentration differs

from equilibrium to non-equilibrium conditions. Thus the relaxation time approximation does not satisfy the condition of current continuity, and is therefore invalid for non-equilibrium conditions.

A new approximation is made to the collision term of the BTE

$$\left(\frac{\partial f}{\partial t}\right)_{\text{COLL}} \simeq \frac{f - g(\vec{k})n}{\tau} \quad (3)$$

$$\int g(\vec{k}) d\vec{k} = 4\pi^3 \quad (4)$$

where $g(\vec{k})$ is a Gaussian function in \vec{k} and n is the carrier concentration. The carrier concentration is given by

$$n = \frac{1}{4\pi^3} \int f d\vec{k} \quad (5)$$

If τ is a function of ordinary space only, then the integral of equation 3 over momentum space is always equal to zero. Therefore this approximation always satisfies the condition of current continuity. Close to equilibrium the new approximation approaches the relaxation time approximation because

$$g(\vec{k})n = f_0(\vec{k}) \quad (6)$$

at equilibrium.

III Discretization in one spatial and one momentum dimension

The discretization of the governing equations in all three spatial and three momentum dimensions would lead to a very large system of equations to be solved simultaneously. Therefore only one spatial and one momentum dimension are included. Poisson's equation is discretized in one spatial dimension using finite differences. The BTE is discretized in one spatial and one momentum dimension using finite differences, the gradients being approximated using central differences. With just one spatial and one momentum dimension it is assumed that the distribution of carriers in the other two momentum dimensions is the same as at equilibrium and that collisions only scatter carriers forwards or backwards in that one momentum dimension. The velocity is modelled on a parabolic band structure. For simplicity the BTE and Poisson's equation are solved using Gummel's method [5] subject to appropriate boundary conditions.

IV Boundary conditions

The contacts are taken to be Ohmic. The potential at the contacts is set equal to the sum of the applied voltage and the built in potential. For the BTE a Dirichlet boundary condition is imposed at the extremities of momentum and a natural boundary condition is applied at the contacts. At the maximum and minimum values of momentum, the distribution function is set equal to the equilibrium distribution function. At the contacts the gradient of the distribution function is set equal to zero assuming there is a neutral space charge region close to the contacts.

V Results

The model is applied to simulating the operation of a lightly doped n^+-n-n^+ diode. The n^+-n-n^+ diode has an overall length of 10 microns with the outer 2.5 microns at each side doped at 10^{14} cm^{-3} . The 5 micron region in the centre is doped at 10^{13} cm^{-3} . The contact at the right is grounded and a bias is applied to the left contact. Figure 1 shows the potential distribution across the device for an applied bias of 400 mV. Figure 2 shows the electron concentration under the same conditions. The distribution function at the right contact is shown in figure 3 along with the distribution function where the carrier concentration is lowest (3.5 μm from the left contact). The greater displacement from equilibrium of the distribution function where the carrier

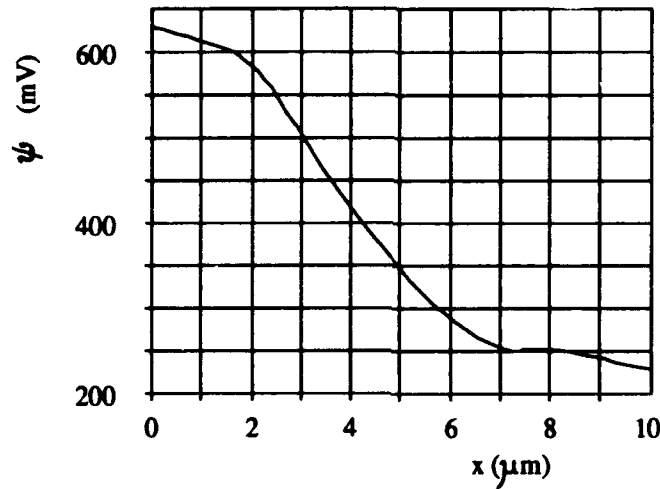


Figure 1: Electric potential (ψ) for a bias of 400 mV applied to the left contact.

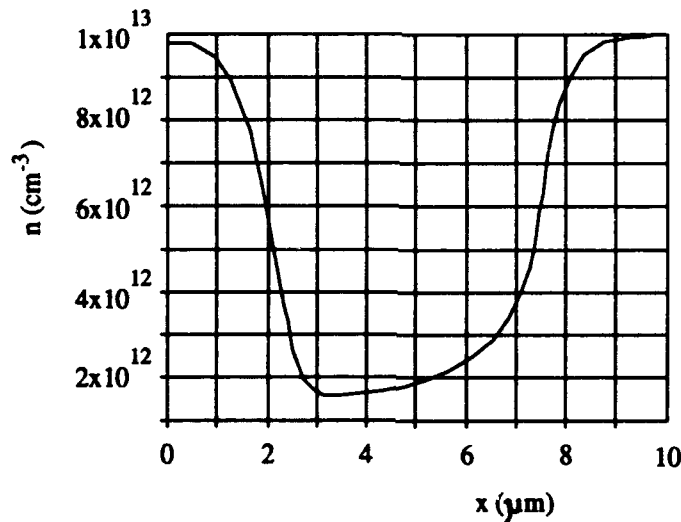


Figure 2: Electron concentration for a bias of 400 mV applied to the left contact.

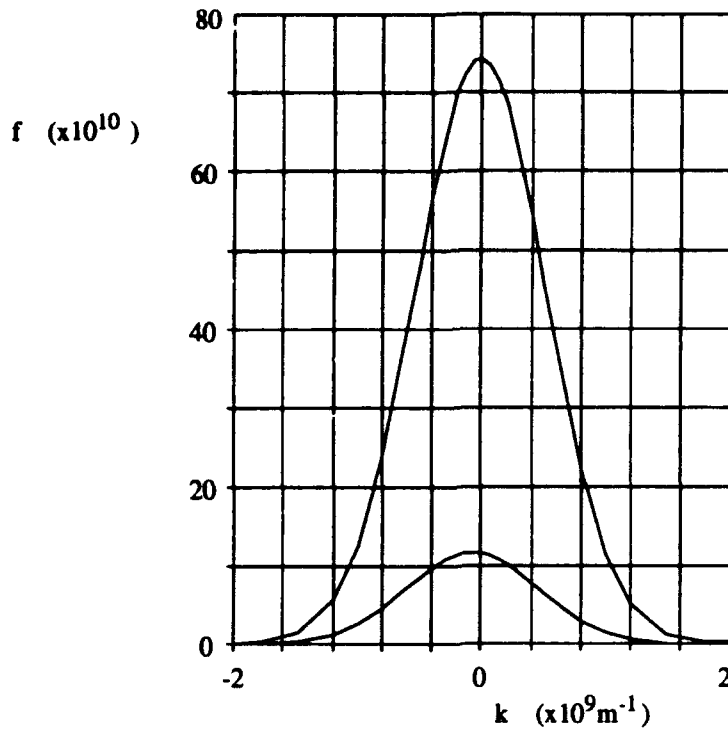


Figure 3: Shape of distribution function

concentration is lower corresponds to a greater average energy of the carriers towards the centre of the device. Figure 4 shows the average energy of electrons at 300K across the device. Note that the carriers a distance 2.5 microns from the right contact (barrier region) have an average energy lower than the lattice thermal energy, a phenomenon which occurs when carrier diffusion is in the same direction as the external force on the carriers [6]. Current continuity is implicit in the model and is observed for all applied biases. Current density in the device is easily calculated from the momentum distribution function. The current density versus applied bias for this structure is shown in figure 5. This result is identical to that predicted by the drift diffusion model.

References

- [1] A. Forghieri, R. Guerrieri, P. Ciampolini, A. Gnudi, M. Rudan and G. Baccarani, "A new discretization strategy of the semiconductor equations comprising momentum and energy balance", IEEE Trans. Computer-Aided Design, Vol.7, No.2, pp.231-242 (1988)
- [2] S.N. Keeney, "The numerical simulation of floating gate non-volatile memory devices", Ph.D. Thesis, p.100 (1992)
- [3] H. Lin, N. Goldsman and I.D. Mayergoyz, "Device modeling by deterministic self-consistent solution of Poisson and Boltzmann Transport Equations", Solid-St. Electron. Vol.35, No.6, pp.769-778 (1992)

- [4] D.E. Dunn and C.A. Paz de Araujo, "GAVELS, a velocity-space GaAs transport model - I. Steady state simulation", Solid-St. Electron. Vol.35, No.4, pp.467-475 (1992)
- [5] S. Selbeherr, "Analysis and simulation of semiconductor devices", Wien New York: Springer-Verlag (1984)
- [6] R. Stratton, "Diffusion of hot and cold electrons in semiconductor barriers", Phys. Rev. Vol.126, No.6, pp.2002-2014 (1962)

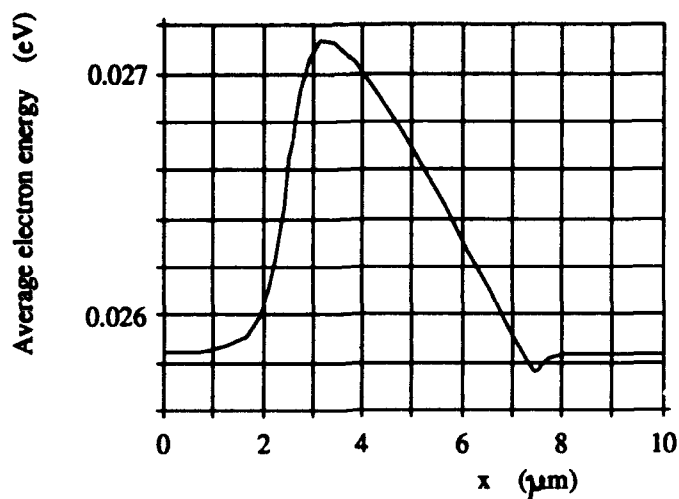


Figure 4: Average electron energy for a bias of 400 mV applied to the left contact.

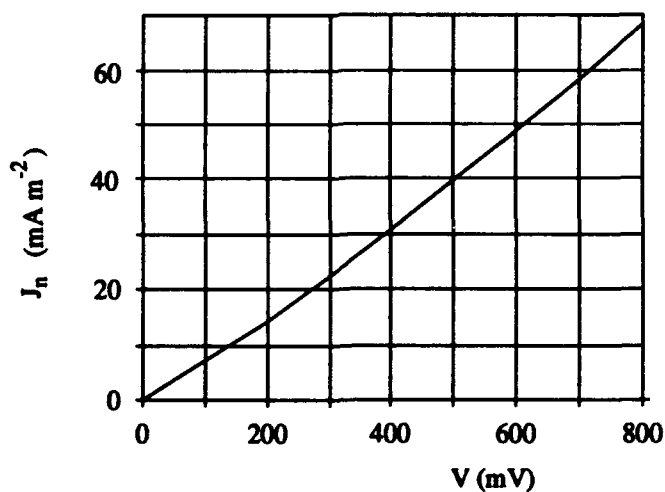


Figure 5: Current density versus applied voltage.

Highly Efficient Simulation of HEMTs and MESFETs Based on Quantum Mechanics

R.K. Veresegyhazy and C.M. Snowden
Department of Electronic and Electrical Engineering
University of Leeds
Leeds, LS2 9JT. UK

Abstract

A new, quasi-two-dimensional physical device model was developed which allows the DC small- and large-signal microwave characteristics of HEMTs and MESFETs to be obtained based on device geometry and process data. It is easily applied to a wide variety of HEMT structures including pHEMT, AlGaAs/GaAs and multichannel structures.

I. Introduction

Field effect transistor physical models can be used in several electronic engineering areas, including device and circuit design. Analytical models, due to their simplicity, can not reliably predict device performance whereas two dimensional models (both hydrodynamic and Monte Carlo) are certainly predictive and usually accurate, but computationally very demanding. Quasi-two-dimensional (Q2D) modelling offers a good compromise as it is up to 1000 times faster than full-two-dimensional modelling and still reasonably accurate. In practice I-V characteristics and microwave S parameters can be obtained in a few minutes on a personal computer with a 80486 processor. Our program can also determine simultaneously the elements of non-linear quasi-static equivalent circuit model. S parameters are calculated using a time-domain version of the physical model.

The Q2D approach is based upon the fact that the equipotential lines in the active channel (i.e. outside the depleted region) of a HEMT or MESFET are fairly parallel, perpendicular to the free surface. Therefore the analysis of the device can be subdivided into establishing the charge control law by the gate in the direction perpendicular to the gate (y direction) and calculating the current transport in the direction parallel to the gate (x direction).

II. Charge Control Law

In the charge control model a vertical cross-section of the device is analysed (considered as a Schottky-diode). The calculation of the sheet electron and ionised donor concentrations are achieved by self-consistently solving Poisson's equation with Fermi statistics and the Schrödinger equation. Fast convergence can be achieved using Newton's algorithm. Up to nine Eigen energies can be evaluated when solving the Schrödinger equation though for practical reasons usually only two or three are calculated, while the rest of the electrons are treated as those belonging to the three-dimensional electron population. When the quantum well is not very deep, as is often the case, it is possible to by-pass the quantum mechanical calculations which makes the simulation even faster.

An advantage of this approach is accuracy and generality, as opposed to the triangular well

or square well approximations, which usually contain an empirical fitting parameter in order to improve accuracy [1].

At the end of calculations a look-up table is produced which contains the gate voltages and the sheet electron and ionised donor concentrations in each epitaxial layer. In case of pseudomorphic structures the effect of strain on the band gap, conduction band edge discontinuity, dielectric constant and effective mass is calculated.

gate width: 1.0 mm

gate length: 0.3 μm

source gate distance: 1.0 μm

gate drain distance: 1.3 μm

layer structure:

AlGaAs(electron supplying layer): 30 nm, Al content: 20%, donor concentration: $1.2 \times 10^{24} \text{m}^{-3}$

AlGaAs (spacer layer): 3 nm, Al content: 20%, donor concentration: $1 \times 10^{20} \text{m}^{-3}$

InGaAs (active layer): 18 nm, In content: 15%, donor concentration: $1 \times 10^{20} \text{m}^{-3}$

GaAs (buffer layer): 150 nm, donor concentration: $1 \times 10^{20} \text{m}^{-3}$

Table 1.: The structure and properties of the simulated pHEMT device

For a pseudomorphic AlGaAs/InGaAs/GaAs HEMT (see Fig.1. and Table 1) the conduction band edge and the electron concentration is shown in Fig.2 and Fig.3 for zero gate bias. The dependence of sheet electron and ionised donor concentration on the voltage applied to the Schottky-diode is shown in Fig.4. It can be seen that at negative gate biases (close to pinch-off) the sheet electron concentration is several orders of magnitude lower than in Fig.3. .

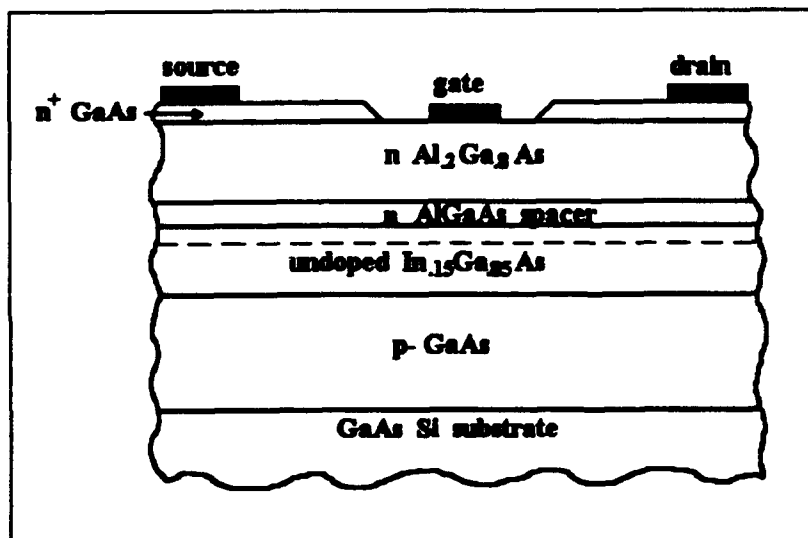


Figure 1. Structure of a typical pseudomorphic HEMT

The quantum mechanical calculations were not by-passed in the present simulation (in favour of Fermi statistics), as it would have led to about 20% overestimation of the sheet electron concentration in the InGaAs layer and 10% overestimation of the total sheet electron density.

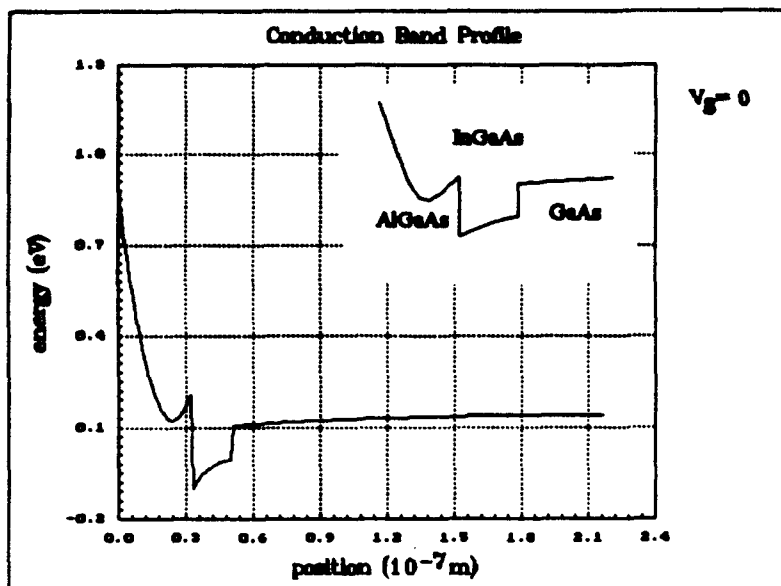


Figure 2. Energy-band diagram of a vertical section below the gate of the pHEMT in Table 1.

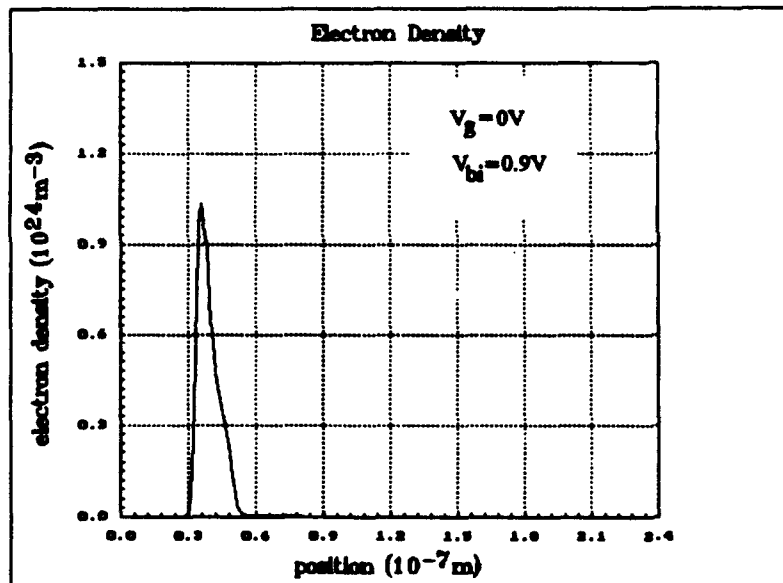


Figure 3. Electron density for the cross-section of a pHEMT corresponding to the conditions shown in Figure 2.

III. Channel Simulator

In the channel simulator Poisson's equation and the transport equations are solved self-consistently in the x-direction assuming the gradient of the Fermi level to be one-dimensional. The transport equations are simplified versions of the first three moments of the Boltzmann equation, the particle (current continuity), momentum and energy conservation equations [2]. The terms describing the scattering of free electrons are calculated based on analytical fits to

Monte Carlo simulation results [3], and are all considered to be the function of average electron energy.

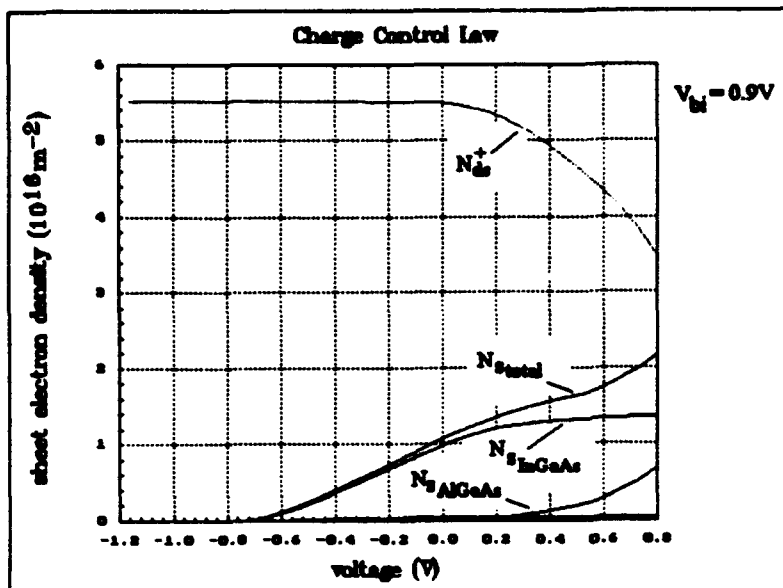


Figure 4. Charge control law: sheet electron concentration and ionised donor concentration in each layer

The sheet electron and ionised donor concentration is found from the above look-up table, based on the difference of channel and gate (or surface) potential, with a correction arising from the gradient of electric field in the x-direction. As this term can be quite large, especially around the drain end of the gate, more accurate calculations should use a double subscripted look-up table for the sheet charge densities to include the $\text{grad}(E_x)$ dependence. Due to the large memory requirement this approach is more suitable for simulations running on workstations than on PCs, therefore an analytic approximation was used in this work.

Unlike other recently published models, our model can simulate parasitic MESFET conduction, allowing a more accurate representation of device operation. In fact, HEMT and MESFET devices with an arbitrary number of different layers can be simulated.

The calculations at a specified source current proceed from source to drain using a simple forward differencing scheme, after omitting the diffusion term in the momentum conservation equation based on the considerations in [2]. The combination of Poisson's equation and the transport equations yields a quadratic equation for the electric field or the drift velocity. The drain voltage corresponds to the channel potential at the drain and is obtained integrating the electric field along the channel (Fig.5). The I-V characteristics of the pHEMT is shown in Fig.6.

Acknowledgement

This work is supported by M/A-COM Inc. Corporate Research and Development Center, Lowell, USA. The authors acknowledge useful discussions with Robert Drury.

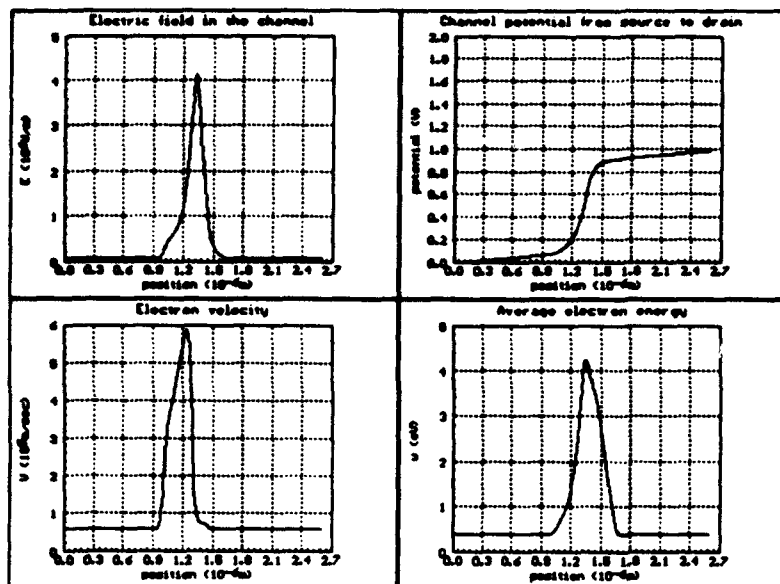


Figure 5. The electric field, channel potential, average electron energy and drift velocity in the conduction channel at bias point: $V_{gs} = -0.4$ V, $V_{ds} = 1.0$ V

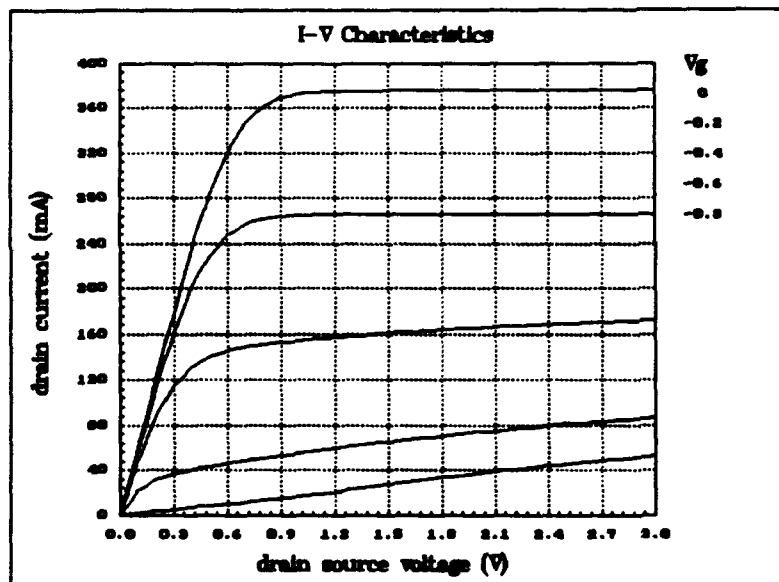


Figure 6. DC characteristics of the simulated pHEMT device

References

- [1] J. Yoshida, IEEE Trans. Electron Devices ED-33, pp. 154-156, 1986
- [2] C.M. Snowden and R.R. Pantoja, IEEE Trans. Electron Devices, ED-36, pp. 1564-1574, 1989
- [3] B. Carnez, A. Cappy, A. Kaszynski, E. Constant, G. Salmer, J. Appl. Phys., 51, pp. 784-790, 1980

The future with computational electronics: a new golden age?

John R Barker

**Nanoelectronics Research Centre
Department of Electronics and Electrical Engineering
University of Glasgow
Glasgow G12 8QQ, UK**

Abstract

The advent of low cost high performance workstations and new visualisation tools is coincident with a growth of new problems in electronics ranging from large scale power devices to nanometer dimensions. The rise of molecular electronics and bioelectronics points to new regimes for the device modeller to explore. It is argued that some software discipline especially standards and portability is required if the modelling community is to be effective. Lessons may be drawn from successes in other fields such as quantum chemistry and molecular modelling.

I. Introduction

Just seven years from the twenty-first century it is perhaps a good time to reflect on the future of computational electronics. The modelling and simulation of solid-state devices is at an important watershed in its history because for the first time it is possible to imagine universal access to software and mathematical techniques which will allow any experimentalist, theorist, industrialist, whatever, access to knowledge which was at one time the select speciality of a few groups around the world. The changes are being brought about by the microelectronics revolution which now provides the power of 1980s mainframes and early supercomputers in relatively cheap workstations - with promises of more to come. There is a similar change in software habits; in the last few years industrial employers of our engineering and science graduates have begun to expect proficiency in languages such as C rather than the traditional FORTRAN; and some nodding acquaintance is now expected with the concept of "user-friendliness". It is no longer fashionable to log up hours of Cray time to produce a couple of numbers. There are expectations aroused that something better is required. Explanations, interpretations, insight, visualisation, projections are required quickly, graphically and easily. Computer scientists are devoting time to topics such as persistent programming which addresses the issues of 25 year old and older software. Computation is now one strand of more complex activities; any computed data is to be regarded as part of a rich interacting hierarchy of

knowledge which can be used by different people for different reasons. In computational electronics we have only just begun to realise what these changes might entail and it is the purpose of this paper to give a personal view of what we have to look forward to. I want to start by first looking back to the beginnings of the subject(for me anyway) in the 1960s.

2. Some history

Coincidentally, it is also a quarter of a century since the first conference on computational electronics held at the Culham laboratories in 1968. At that meeting were some of the first papers on computational electron transport theory including one by myself and Cliff Hearn which dealt with self-scattering in a Monte Carlo simulation of photo excited hot electrons and which used propagator techniques to prove that the Monte Carlo solutions were equivalent to solutions of the Boltzmann equation. At that time Culham was a centre for computational work, mainly due to the intense interests in plasma physics and to a lesser extent neutron transport theory. Culham provided an educational distribution point with series of weekly lectures on computational physics organised by Potter. The smallness of the computers of those days made the use of compact ingenious algorithms the order of the day rather than brute force number crunching. It is not surprising that Potter went on to develop the Psion organisers and other compact systems.

By 1969 much of the familiar theoretical base of computational electronics was already in place. JFET theory had been set up in 1947, Bipolar transistor theory by 1952, MOSFET theory by 1960, MESFETs in 1966. Of the computational techniques, drift-diffusion theories dated from 1952, hydrodynamic models from 1962, the Gummel algorithms from 1964 and the first solid state electron transport theory handled by Monte Carlo was reported by Kurosawa in 1966(the latter work was part of the inspiration that guided me into transport theory especially hot electron theory in 1967). But many of these techniques were really derivative: Monte Carlo studies were enjoying extensive use in neutron transport theory (although Lord Rayleigh was doing much the same thing without computers in the last century); and the influence of hydrodynamics (particularly driving the aerospace industries from WWII onwards) and plasma physics cannot be under-estimated. There were already major forays into quantum kinetic theories although these were still of a formal rather than practical nature.

What really limited practical applications was computing power. The really interesting (as well as applicable) problems were either highly complex or non-linear or both. Looking back I am still amazed by the lack of imagination we all had. "You are only limited by your imagination" is the proud boast of many a computer advertisement in the 1990s. Well, my limit was a factor between 3 and 10. In 1967 I developed code (ALGOL 60) for simulating the time evolution of Gunn domains and similar effects in photoconductors using an Elliot 803 computer with 32K

core memory. A little later the offer of time on the national ATLAS computer at Harwell brought the luxury of 90K memory but entailed a two day wait to receive the wad of data (mainly error messages and core dumps) from the post. That extra factor of 3 in memory seemed impressive, and every few years we in the computational sciences looked forward to that extra factor of ten or so which would bring the biggest of problems into our control. It is perhaps fortunate that the compute did not exist because the control and detailed understanding of solid state device materials was not sufficient for realistic modelling. Not only was compute power singularly poor, but means of handling the output data were primitive. In 1968 I spent maybe one week at a time collecting plotter pictures on large rolls, selective frames of which were placed along a corridor to facilitate a "computer movie" of Gunn domain dynamics. Today, that entire problem, including symbolic manipulation and computer movies can be handled using Mathematica on a Macintosh computer in a few minutes. It is this simple reminiscence which brings home the immense opportunities that we now have for unleashing computational electronics from the grip of computational poverty. But others have been there first: the computational chemists have achieved a close partnership with their industry which has produced an effective, widely shared approach to applicable computational science which we in computational electronics could learn from.

III. Physical modelling versus equivalent circuit

There has always been a division between those advocating the empirical approach to device and particularly circuit modelling and those favouring the development of physical models. The advantages and limitations of the two approaches are well known: industry, especially given the exponential pace of integrated circuit technology, has needed data quickly and cost-effectively; empirical or equivalent circuit models do the trick - in the short-term. But, that approach and a poor investment in physical models led one well-known review paper in the late 1970s to state that it was highly unlikely that silicon transistors could be manufactured with design rules less than 10 microns. It is also true that physical models give the best route to device understanding and to devising new device concepts - provided the underlying physical model is reliable. The latter case is still a problem because we do not usually know how to model surface and interface charge very readily. Indeed, the whole problem of physical modelling of real devices is bound up with understanding and developing models of the underlying technological processes.

The physical modellers have often dealt with the esoteric: the spate of quantum "devices" investigated in the 1980s would have achieved short shrift with the empiricists had they followed the literature. After all it is assumed that devices should be interconnectible, isolatable and designed to function at room temperature not milli-degrees Kelvin. Again due to a certain

innocent over-familiarity with the field many advocates of quantum devices fail to point out the frequent necessity to use lock-in amplifiers in operating the devices. The real lesson here is not that work on quantum devices is only of academic interest: it is that the science required to achieve engineering goals is not necessarily the same science that drives the physical scientist to investigate phenomena within devices. This problem did not arise with the quantum chemists. Over the last 25 years the chemistry community has mutually developed molecular modelling, drug design and molecular graphics/design software which is not only relevant to science and industry but mixes the empirical and physical to advantage; is also readily available and widely used. My own interests in molecular electronics have led me to have a healthy respect for the sheer accessibility of computational techniques and databases to the chemistry community.

There are distinct signs now that the physical modelling/empirical industrial rift is healing within the electronics community. This is coming about partly because of better control and understanding of real materials and the need to include "esoterics" in device design e.g. hot electron effects. There are also questions of "attitude": we have all heard of the physical scientists easy dismissal of engineering problems as containing no interesting physics. Often this is misleading and in our own recent work (with Asenov and Brown) on applying (technology transfer ?) microscopic device modelling techniques to large power transistors (200 A switches) we have developed considerable respect for the complexity and subtlety of the underlying physical problem. More significantly, the immense complexity of realistic quantitative as opposed to qualitative physical device simulation such as Laux's Monte Carlo models of MOSFET devices is becoming manageable with the advent of very high power workstations.

IV. Machine power I: workstations

Early next year my University will relinquish its three mainframes (including an IBM 3090 vector processor) in favour of a distributed system of workstations inter-connected by fast networking. Other Universities here and abroad have already experienced this trauma - and survived. The argument is compelling: why share a 126 Mbyte mainframe with 400 other people when you can have a cheap 256 Mbyte workstation, with instant on-line colour graphics for yourself. For a little more money one can expect to have 100 MFLOP workstations for use within modelling groups over the next few years. If the dreams of modellers from the 1970s are to be believed this already is extreme overkill for computational electronics problems. Of course that is not the case, but it is the case that this development should lead to very much closer links between academics and industry on grounds of cost alone.

There are questions of fashion however. There is still some cachet in using SUN workstations. My own pedestrian approach is to use networked Apple Quadras running Unix,

X-Windows and Mac environments (and even DOS!) concurrently, which at lower cost and some performance penalties allows everyday word-processing, graphics, movie making, e-mail and so on to be integrated with the normal workstation role. The catch of course is that the system is linked to a large transputer array for compute-intensive tasks. This example raises the issues of hardware and operating system standards. My own solution is partly cost based (more for less) and partly with an eye on upgrade routes and inter-working with IBM systems, UNIX, DOS and MAC. It also raises the issue of the human interface: should we be developing specialist code or code that is easily used within familiar environments by non-specialists? If the answer is yes we have to confront the problem of huge software design costs and that extra memory and power starts to look less effective.

What power do we really need? As we shall discuss later some problems require GFLOP -> TFLOP power and > GBytes memory. How do we get it? National super computers or local solutions?

V. Machine power II: vectorisation and parallelisation

Even with the projected advances in integrated circuits it is unlikely that desktop single processor GFLOP power is imminent. Parallel processing in the widest sense of the word is usually advocated as the way ahead and of course forms the basis of many super computer approaches. Vectorisation is the least flexible in one sense and so is easier to use, full-scale parallelisation, for example in a software configurable transputer array is often highly efficient in run-time but may be costly in development time. If the new found universality offered to computationalists by the workstation revolution is not to be over-turned it is essential that some order and standardisation is brought into parallelised codes for general and local use. Our own approach at Glasgow is to cut development time and improve portability by re-generating all our codes in common parallelisable format, for which we have chosen finite-element algorithms as the main format which can cover classical device modelling through to many electron quantum transport. These codes are designed to run on arbitrary numbers of transputers and so can be developed in simple workstations before transfer to the main cluster. It appears at present that the main lesson is to use the workstation as the main hardware tool but supplemented by networking to compute intensive local parallel machines of intermediate power.

These arguments for the workstation approach pale into insignificance when we consider the opportunities that workstations now offer for visualisation.

6. Visualisation

The chemistry community was quick to develop visualisation tools, many of which are now available in the public domain for quite ordinary workstations. It is thus possible to put on appropriate goggles (plus IR sensor) and guide the viewer through a colour stereo journey through a molecular structure on a Silicon Graphics workstation. Virtual reality systems have been pioneered by the computational chemists. In a sense this is obvious, the complexity of molecular design is in direct space: it is three-dimensionally geometrical and topological. In device modelling the effective parameter spaces are less physically obvious and often high dimensional. We have yet to learn how best to represent dynamic data for real devices. The 3D representation of electron and hole trajectories colour coded for temperature is already good practice and drawing a leaf from quantum chemistry the use of 3D transparent (cloud representation) colour contours for potentials of temperature surfaces should be manageable. This requires 24 bit colour high resolution workstations and of course soaks up memory and CPU time (graphics accelerators of course are a way out). At Glasgow we have been developing movie and graphics techniques to help us gain intuition into the transient few electron problem in quantum waveguides and single electronic systems. Like the mediaeval schoolmen there is nothing to be gained by not trying things out. I believe that visualisation is crucial to the development of computational electronics especially if the aim is to provide access to general designers and experimenters. Multiple displays with friendly interfaces will be essential. This route should not be left to chance it should be organised and draw on the experience of other communities. In particular the input/output of visualisation should be linked to other software such as CAD codes where possible.

VII. Software

The demands of visualisation, operating systems, portability, parallelisation bring us to the prospects for the software side of the business. It also raises the issue of what level we should develop our computational models. First some generalities. It is becoming obvious that computational electronics involves large codes which are both continuously developing and meeting new demands from interface and visualisation requirements. Although there is huge investment in it, FORTRAN is not the best medium for managing and verifying large quickly developable, de-buggable codes (my own bias was already admitted earlier - I was brought up on ALGOL). C, C++, PASCAL all have their merits especially in object oriented forms. The answer is not to choose the best (I actually am attracted to functional programming languages) but to choose the best for the community.

On visualisation should we be developing our own codes or providing hooks into public domain or commercial software? I would argue for both because today's best visualisation software has evolved from earlier scientific studies.

One of the most interesting developments has been the spread of software such as MATHCAD and MATHEMATICA which provide highly simplified frameworks for quickly solving what used to be formidable problems (with excellent graphics thrown in). The use of symbolic techniques and functional programming make these approaches very attractive for certain problems (for example simple 1D tunnelling simulations). But they have little efficiency and it is often difficult to program them for efficiency. Nevertheless these approaches do provide universality.

VIII. Organised modelling

Like synchronised swimming I believe that organised modelling will eventually appear sensible and exciting. We need to learn from communities like the chemists that there is mutual advantage in sharing, documenting and jointly developing community codes for device modelling. There are too many stand-alone operations to bring the promise of universal access to modelling closer.

There is a great deal of in-house expertise which is fiercely guarded (not always with justification). I would like to see this type of workshop develop into a forum that aims to set standards, goals, achieve inter-working and sharing of knowledge so that full attention can be paid to applications driven research. Paradoxically some organisation is necessary because of the problems of diversification which come from the freedom brought about by the workstation revolution. How this should be achieved is a subject for debate.

IX. New opportunities

Finally, let us turn to a selection of the new opportunities for the core of computational electronics - the models, applications and tools themselves. Many of these are represented at this workshop, some are steeped in antiquity, others are striving for birth.

In classical/semi-classical models the challenges are in TeraHertz high frequency device modelling, the inclusion of thermal effects, the incorporation of process models. Despite the complexity it is becoming possible to devise physical models which quantitatively describe real devices. These will be essential as integrated circuit technology heads into the fully sub-micron regime. The advent of surface probe technology (STM, AFM etc.) and on-chip electrometry and capacitance techniques will open up the hitherto problematic surface and interface charging problems which will underpin the model development. There are considerable opportunities for modelling which is designed to interpret the new atomic scale probes.

The same on-chip techniques have recently allowed direct observation of the ubiquitous fluctuation potential which bedevils quantum devices and the direct observation of many

electron states and their filling in quantum dot structures. The application to ultra-small commercially oriented devices will be challenging and will require active modelling work. The aim is to understand and control the intra- and extra-device environment in detail.

The shift to ultra-small devices and devices with ultra-small capacitive regions brings up the issue of quantum charging effects: Coulomb blockade, correlated tunnelling, macroscopic quantum tunneling, telegraph and shot noise effects associated with the discreteness of the electronic charge. The modelling required here and for the now extensive range of single electronic systems is non-trivial. Single electronics is scaleable to high temperatures by fabricating sufficiently small capacitive tunnel junctions. It is already integratable. Techniques from the charge density matrix, the quantum Langevin equation, Monte Carlo, Traffic theory and linear programming are all being brought to bear. The single electron problem actually turns out to be a many body problem and much remains to be discovered. Device-device coupling is an unavoidable reality in such systems and extended systems of devices need to be modelled. Control of killer effects such as quantum fluctuations, charge-trapping de-trapping and cross-talk can in principle be handled by designing the substrate electrostatic environment correctly: 3D modelling on heterostructures is essential here.

There is now enormous scope for quantitative quantum modelling but we should be careful to distinguish the physically interesting from the engineering possibilities. To restore some balance I should remark that my own interests partly involve studying quantum device systems fabricated within two-dimensional electron gases which can act as atomic scale instrumentation for exploring hitherto intractable regimes of quantum physics such as delayed-choice, empty wave and quantum measurement problems. In this context we have been recently successful in modifying the Bohm pilot wave version of quantum mechanics in setting up and solving problems in quantum transport theory.

If we pursue the limit of miniaturisation of devices to its logical conclusion we encounter structures on molecular scales. The problems of bioelectronics, biosensors, molecular sensors, molecular electronics and nanoelectrochemistry all converge with device electronics at nanometre scales. Even the techniques in quantum chemistry for studying the STM tunnel current and energy spectrum of a small molecule bound to a surface bear striking resemblance to quantum dot modelling. The long range tunnelling of electrons through polymer chains, redox centre arrays and enzymes is closely related to the electronic soliton dynamics of single electronic systems.

The future then is rich in promise, quantitative models, evolving interdisciplinarity and a chance to develop a strong mature modelling community properly integrated with industry. I look forward to a new golden age.

Discretisation of the Hydrodynamic Transport equations using the Control Region Approximation

P.A.Mawby and A.McCowen
Department of Electrical and Electronic Engineering
University College of Swansea
Singleton Park
Swansea SA2 8PP

Abstract *This paper focuses on the discretisation of the hydrodynamic transport equations for electrons in sub-micron MOSFET structures. In particular initialisation of the electron temperature equation, and the important problem of accurately calculating the energy input term ($J \cdot E$) on a generalised triangular FE mesh will be considered. The discretisation is applied to the calculation of substrate leakage currents in sub-micron LDD MOSFETs, the results of which will be presented.*

I. Introduction

As the size of active devices in VLSI sub-systems is continually reduced the validity of the standard drift-diffusion equations, as commonly applied to model semiconductor devices, is brought into question. The basic assumption made in the derivation of the drift-diffusion framework is that the charged carriers in a device are in thermal equilibrium with the lattice phonons at some ambient temperature.

The kinetic Boltzmann transport equation is the starting point for describing a many-particle carrier-phonon system, however, direct solution of this equation is a massive computational task and is therefore generally not attempted. A more practical approach is to use the first-five moments of the Boltzmann transport equation to provide an approximate solution which neglects the higher-order moments. These moment equations are commonly referred to as the hydrodynamic equations, which are basically conservation of charge, conservation of momentum (one in each of the three spatial dimensions) and the conservation of energy. This is an acceptable approximation provided there is sufficient randomisation of the thermal energy in the system to allow the meaningful use of average quantities such as velocity and effective carrier temperature. Under these conditions, the hydrodynamic equations, when used in the relaxation time limit, provide a practical engineering platform, which can be used to investigate hot-carrier effects such as velocity saturation and velocity overshoot.

This paper describes the discretisation of these equations using the Control Region Approximation. This method is general and is not restricted to a particular shape of element, however, we will assume the use of triangular elements. This has particular importance in the definition of the energy input into the energy moment equation. This will be discussed more fully in section 6.

II. Physical Equation System

Application of the moment method to the Boltzmann Transport Equation produces a series of equations, one for each moment taken. The resulting steady state electron energy moment equation can be written as

$$\nabla \cdot \mathbf{S}_n - \mathbf{E} \cdot \mathbf{J}_n + R w_n + n \frac{w_n - w_e}{\tau_{en}} = 0 \quad (1)$$

here \mathbf{S}_n is the energy flux vector, the second term describes the energy input into the system, and the last two terms describe energy loss in the system the first due to carriers recombining and the last due to inelastic collisions in the device. The average electron energy is described by w_n , R is the recombination rate and τ_{en} is the energy relaxation time. The energy flux term is described by the constitutive relationship

$$\mathbf{S}_n = \mathbf{Q}_n - \frac{\mathbf{J}_n}{q} (w_n + kT_E) \quad (2)$$

here \mathbf{Q}_n is a heat flux term which is identically zero under the symmetric Maxwellian assumption, however it is included heuristically here to account for the energy flow due to carrier thermal gradients.

The conservation of carriers is described by the steady state continuity equations for electrons and holes

$$\nabla \cdot \mathbf{J}_n - qR = 0 \quad (3)$$

$$\nabla \cdot \mathbf{J}_p + qR = 0 \quad (4)$$

which again require the constitutive relationships for flux in this case current densities. For electrons the appropriate moment equation along with the assumption that the momentum relaxation time is very short compare to the energy relaxation time leads to

$$\mathbf{J}_n = -q\mu_n n \nabla \psi + \mu_n k n \nabla T_E + \mu_n k T_E \nabla n \quad (5)$$

where μ_n is the electron mobility, n is the electron concentration, ψ is the electrostatic potential, k is the Boltzmann constant and T_E is the effective electron temperature. Hole transport is assumed to be described by the standard Drift-Diffusion framework.

$$\mathbf{J}_p = -q\mu_p p \nabla \psi + \mu_p k T_L \nabla p \quad (6)$$

where T_L is the ambient lattice temperature. Finally, Poisson's equation is solved for electrostatic potential

$$\nabla \cdot \mathbf{D} - \rho = 0 \quad (7)$$

where \mathbf{D} is the electric flux vector given by

$$\mathbf{D} = \epsilon \mathbf{E} = -\epsilon \nabla \psi \quad (8)$$

ϵ is the dielectric constant for the material. Equations (1), (3), (4) and (7) along with the constitutive relationships (2), (5), (6) and (8) form a closed set of equations in the variables T_E , n , p and ψ . Which can be discretised using the control region approximation.

III Physical Models

Mobility model

At elevated electron temperatures carrier mobility is reduced by increased scattering rates this causes the carrier velocity to saturate. The electron temperature can now be used as a parameter in the electron mobility model. In effect it plays the same role as the electric field does in the standard drift-diffusion context. In this work an electron temperature dependent model is used which is consistent with the conventional field dependent model for mobility under homogeneous conditions, which has the form

$$\mu_n(T_E) = \mu_o \left\{ 1 + \lambda \left[\frac{k_B T_E}{q} - \frac{k_B T_L}{q} \right] \right\}^{-1} \quad (9)$$

μ_o is the low-field mobility and λ is related to the electron saturation velocity v_s and the energy relaxation time τ_{en} by

$$\lambda = \frac{3}{2} \frac{\mu_o}{v_s^2 \tau_{en}} \quad (10)$$

Impact Ionisation Model

In hydrodynamic simulations, non-local effects on carrier transport are of primary concern. Energy consideration in carrier transport plays a crucial role at this level of simulation. Accordingly, the suitability of using local field-dependent impact ionisation models in hydrodynamic simulations is not at all obvious. To alleviate this problem, the electric field dependence of the well-known Chynoweth's empirical formula for impact ionisation coefficient α_n is replaced by an appropriate electron temperature dependence. This requires the establishment of a relationship between the electric field and electron temperature. For the homogeneous case, the energy equation (1) together with the mobility model (9) can be used to obtain such a relationship

$$E = \left\{ \frac{3}{2} \frac{k(T_E - T_L)}{q \tau_{en} \mu_o} \left[1 + \frac{\lambda k}{q} (T_E - T_L) \right] \right\}^{1/2} \quad (11)$$

Using the Chenoweth formula for α_n

$$\alpha_s = a \exp\left(-\frac{b}{|E|}\right) \quad (12)$$

where a and b are empirical coefficients, one can model the non-local effects of impact ionisation through the non-local solution of electron temperature.

IV Control Region Approximation

The partial differential equations that make up the device model are all in the divergence form, that is they have the divergence of some flux vector quantity and a number of source terms. In general then they can be written

$$\nabla \cdot \mathbf{F} - S(\mathbf{r}) = 0 \quad (13)$$

Over some local region of the device Ω performing a surface integral gives

$$\begin{aligned} \iint_{\Omega} \nabla \cdot \mathbf{F} d\Omega - \iint_{\Omega} S(\mathbf{r}) d\Omega &= 0 \\ \Downarrow \\ \oint_{\Gamma} \hat{n} \cdot \mathbf{F} d\Gamma - \iint_{\Omega} S(\mathbf{r}) d\Omega &= 0 \end{aligned} \quad (14)$$

where \hat{n} is the outward normal unit vector and Γ is the boundary of the region. This is now in a form suitable for discretisation. Consider a typical node i in a triangular mesh as shown in Figure 1.

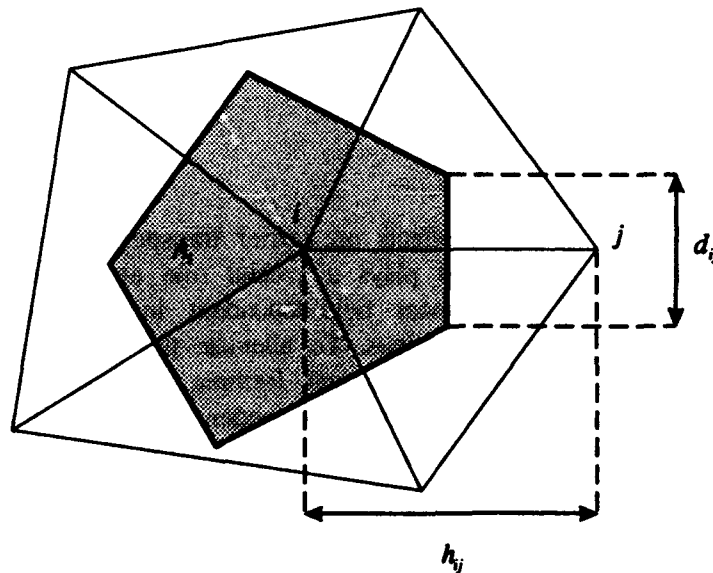


Figure 1. A control region Ω (shaded area), of area A_i , surrounding node i

The region of integration Ω becomes the Voronoi region (shaded) and the boundary Γ becomes the boundary of the Voronoi region. The line integral in (9) can now be approximated by a piece-

wise summation of the flux across each of the edge segments. The integration of the source term is approximated by assuming it to be constant at the value found at the actual node, thus

$$\sum_{i=1}^{M_i} F_{ij} d_{ij} - S_i A_i \quad (15)$$

A_i is the Voronoi area associated with node i , M_i is the number of edge connected to the node, and d_{ij} is the pipe width connecting the nodes along line ij . So providing we can produce a suitable method of discretising the flux vectors along an edge we can apply (11) to all nodes in the mesh, the resulting set of non-linear equations can be solved using the Newton-Raphson method.

V Discretised form of the Semiconductor Equations

The set of divergence equations to solve are

$$\sum_{i=1}^{M_i} S_{ij} d_{ij} - \left(\mathbf{E} \cdot \mathbf{J}_n + R w_n + n \frac{w_n - w_p}{\tau_{wn}} \right)_i A_i = 0 \quad (16)$$

$$\sum_{i=1}^{M_i} J_{nij} d_{ij} - R_i A_i = 0 \quad (17)$$

$$\sum_{i=1}^{M_i} J_{pij} d_{ij} + R_i A_i = 0 \quad (18)$$

$$\sum_{i=1}^{M_i} D_{ij} d_{ij} - \rho_i A_i = 0 \quad (19)$$

The constituent relationships are readily discretised, for the first three equations the modified Scharfetter-Gummel method is used, and for the electric flux the standard finite-difference expression is used. The equations are assembled using the standard finite element approach on an element by element basis. The linear equations are solved using the Bi-CGSTAB method [1].

VI Discretisation of the Energy Source Term ($\mathbf{E} \cdot \mathbf{J}_n$)

This term provides energy input into the electron ensemble, and its proper discretisation is therefore essential. The difficulty in discretising this term arises from the inconsistencies of edge currents in a element. When the electric field (\mathbf{E}) is to be determined, a unique value for an element is readily found from the nodal potentials. This, however, this is not true for the current density field (\mathbf{J}_n), since the Scharfetter-Gummel method used in the discretisation will only yield consistent edge currents when the current flow is purely one-dimensional. In order to resolve this problem several techniques have been suggested [2,3,4]. The first method of Laux is found to be extremely unstable, especially in the drain depletion region, where there are rapid changes in carrier concentration and the weighting method can yield a distribution of heating which is both erratic and unrealistic. The other two methods use completely different approaches to the discretisation of the ($\mathbf{E} \cdot \mathbf{J}_n$) term, nevertheless, the discretised forms are almost identical apart from small second order differences. These terms have significant effect on the stability of the

solution procedure, since they occur in off diagonal terms of the system Jacobian, and under some conditions destroy the diagonal dominance of the system. This means that unless an extremely small mesh spacing is used, which is unphysical and computationally expensive, then a solution cannot be found for drain bias values over a few volts. For this reason the power scheme [3] is preferred, as it does not present any apparent convergence problems.

VII Solution algorithm

Figure 2 shows a flow diagram of the solution algorithm. Initially a fully coupled drift-diffusion solution is sought which is then used to provide an initial guess for electron temperature. It also provides a good guess for the other solution variables. A decoupled scheme is then used to obtain a self consistent solution of the hydrodynamic equations. Typically the number of iterations required for self-consistency increase with an increase in electron temperature.

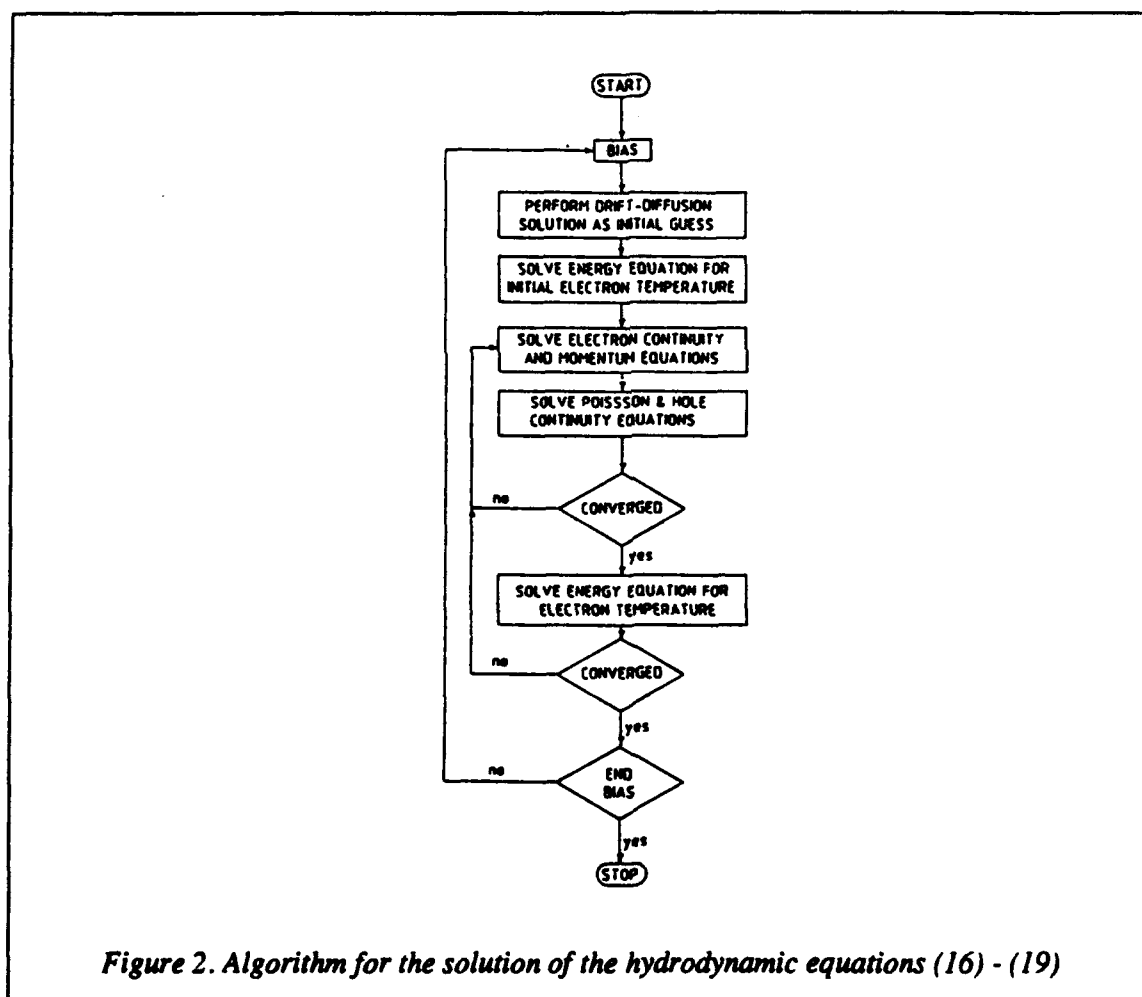


Figure 2. Algorithm for the solution of the hydrodynamic equations (16) - (19)

VIII Results

The results of the hydrodynamic simulation are demonstrated for a LDD device structure, and validated against experimental data on a range of devices of differing gate lengths. Figure 3 shows the electron concentration for a 5 μ m long channel device and the associated electron temperature. As expected the electron temperature is highest at the drain depletion edge, there is also a smaller amount of heating in the channel depletion region. Once the temperature distribution is established at a particular bias point the local impact ionisation rate can be calculated bas on the expression given in equations (11) and (12), from which the substrate current can be evaluated by summation across the device and weighting by the Voronoi area

$$I_{\text{sub}} = \sum_{\text{all nodes}} qnv_s \alpha_{\text{si}} A_i \quad (20)$$

where v_s is the saturation velocity. Figure 4 shows the terminal currents for a 0.8 μ m length device, compared with experimental results for the same structure. In order to get such a good fit an energy relaxation time of 0.02ps has been used, and a saturation velocity of 9×10^6 cm/s has been used. The agreement with these same parameters is equally as good for shorter gate lengths.

IX Discussion

In his paper we have presented a self-consistent model capable of solving the hydrodynamic equations in semiconductor devices. The equations have been discretised on a general mesh using the control region approximation, careful consideration has been paid to the discretisation of the heat source term.

Results of the method have been demonstrated for an LDD MOSFET structure and found to give excellent agreement with measure drain and substrate currents.

REFERENCES

- [1] "Bi-CGSTAB: A Fast and smoothly converging variant of Bi-CG for the solution of non-symmetric linear equation systems" SIAM J. Sci. Stat. Comput. Vol. 13, No. 2, pp631-644 (1992)
- [2] S.E.Laux and R.G. Byrnes ' Semiconductor device simulation using generalised mobility models'. IBM J. Res. Develop., 19, pp. 289-301 (1985)
- [3] A. Gnudi and F.Oden, 'An efficient discretisation scheme for energy-continuity equations in semiconductors'. Simulation of Semiconductor Devices and Processes, Vol. 3., G. Baccarani and M. Rudan (Eds), University of Bologna, 1988 pp. 387-389.
- [4] K.W.Chai, P.A. Mawby and A. McCowen 'Hydrodynamic Simulation of Electron Heating in Conventional and Lightly Doped Drain MOSFETs with application to Substrate current calculation' Int. J. Num. Modelling, Electronic Networks, Devices and Fields, Vol. 5, No. 1, pp.53-66 (1992)

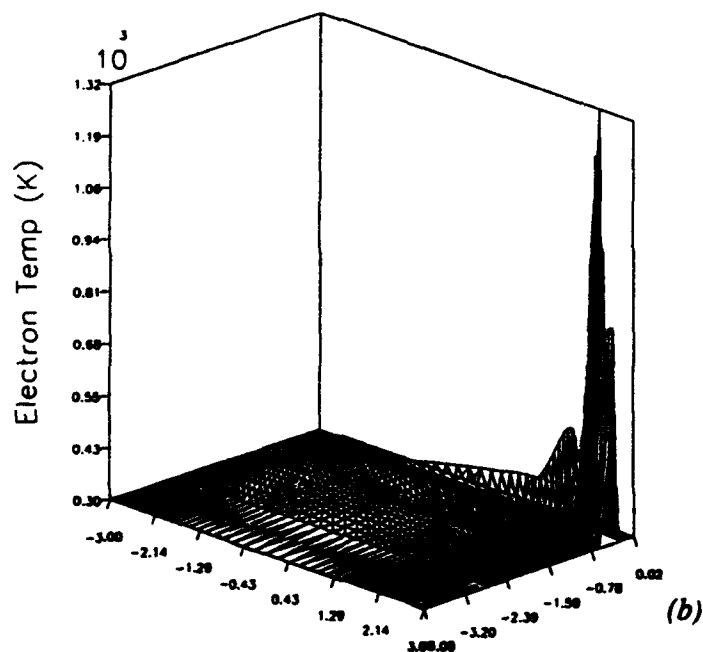
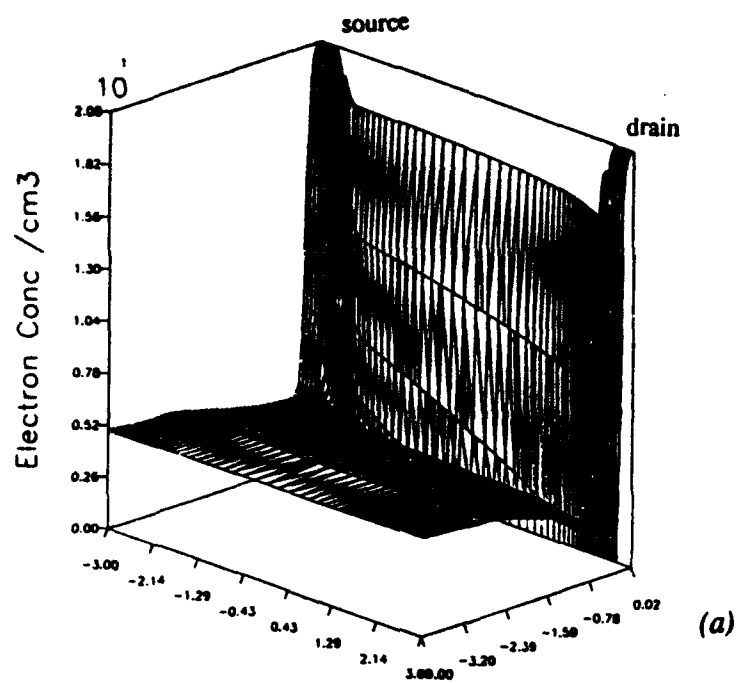


Figure 3. (a) electron distribution $V_{DS} = 3.0V$, $V_{GS} = 3.5V$
 (b) electron temperature $V_{DS} = 3.0V$, $V_{GS} = 2.5V$

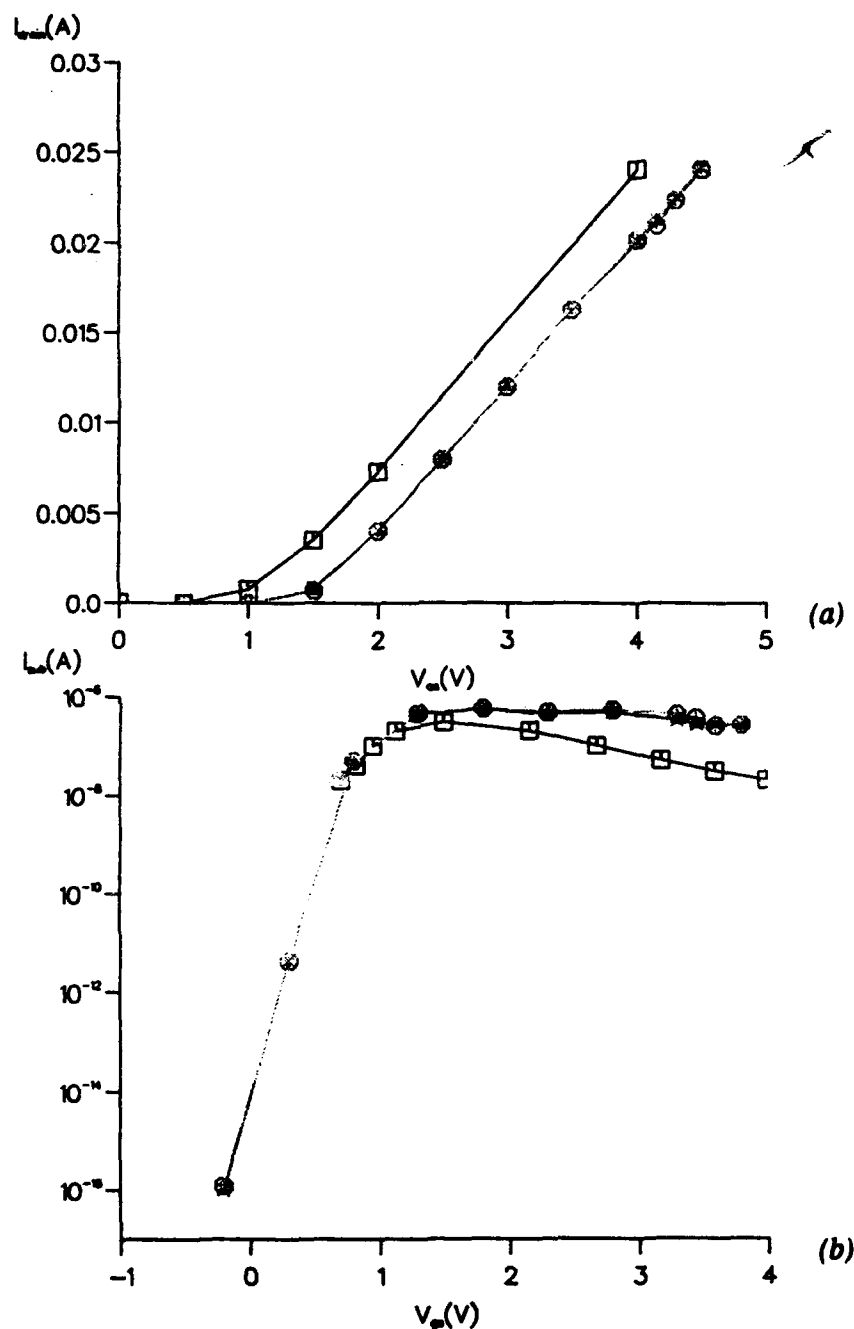


Figure 4. Results from the hydrodynamic simulator

(a) drain current - experimental (□) and simulated (○)
 (b) substrate current - experimental (□) and simulated (○)

2-Dimensional Solution to the Boltzmann Transport Equation to Arbitrarily High-Order Accuracy

Ken Hennacy, Neil Goldsman and I D. Mayergoyz

Department of Electrical Engineering,
University of Maryland,
College Park, MD 20742

Abstract

In this work, we present a general 2-dimensional spherical harmonic formulation of Boltzmann's transport equation. Until recently, numerical implementation of this approach has been discussed for either 1-dimensional geometries, or only a few of the spherical harmonics [1, 2]. In this paper, a formulation is presented that includes an arbitrary number of spherical harmonics.

I. Introduction

Device modeling by direct solution to the Boltzmann transport equation is usually not performed because of dimensionality problems and difficulties in evaluating the collision integral. To directly account for 2-dimensional device operation with the Boltzmann equation, we would normally have to perform calculations in 5 dimensions (2 dimensions in real-space and 3 dimensions in momentum-space). To overcome this 'curse of dimensionality' a new approach to solve the BTE in device models is being developed which uses a spherical harmonic (SH) or a Legendre polynomial expansion.

II. Indefinite Spherical Harmonics Expansion

One of the reasons why we are interested in this method is that it gives us the ability to produce differential-difference operators for the evaluation of the collision integral to all orders in the expansion. Finding expressions for the rest of the BTE operators, however, involves some work. To solve the 2-D BTE to high-order accuracy, the distribution function is expressed in terms of an infinite spherical harmonic expansion with unknown coefficients that depend on energy and position:

$$f(\mathbf{r}, \mathbf{k}) = \sum_{lm} f_l^m(\mathbf{r}, \epsilon) Y_l^m(\theta, \phi) \quad (1)$$

$Y_l^m(\theta, \phi)$ are the spherical harmonics[3]; $f_l^m(\mathbf{r}, \epsilon)$ are the coefficients which are to be determined; $l = 0, 1, 2, \dots$; and for each l , the superscript $m = -l, -l+1, \dots, 0, \dots, l-1, l$.

The spherical harmonics give the angular dependence of the distribution function in momentum space, and the coefficients provide its magnitude. The SH-numerical formulation allows us to account for the angular dependence of the distribution function in momentum space (θ, ϕ) analytically, thereby reducing the dimensionality of our calculations from 5 to 3.

III. SH Formulation to Arbitrarily High Order Accuracy

The objective is now to determine the unknown coefficients $f_l^m(\mathbf{r}, \varepsilon)$, which can be used to construct the distribution function. Furthermore, to minimize the possibility of truncation errors in the SH approach, we have developed a technique to determine the coefficients to arbitrarily high order. The basic idea behind the approach is to automatically generate a system of equations for all the unknown SH coefficients $f_l^m(\mathbf{r}, \varepsilon)$, and then solve the system and construct the distribution function using the above summation.

To generate this set of equations, we first substitute the above summation into the BTE. Next, we project the BTE onto each of the SH basis functions. The projection onto the l, m 'th SH basis function, which yields an equation for the l, m 'th coefficient, is illustrated by the following operation,

$$\int d\Omega Y_l^{m*}(\theta, \phi) \left\{ \left(\frac{1}{\hbar} \nabla_k \varepsilon \cdot \nabla_r - \frac{e}{\hbar} E(\mathbf{r}) \cdot \nabla_k - \left[\frac{\partial}{\partial t} \right]_c \right) f_l^m(\mathbf{r}, \varepsilon) Y_l^m(\theta, \phi) \right\} = 0 .$$

By performing a similar projection onto each of the SH basis functions, the angular dependence of the distribution function is integrated out, and an infinite system of coupled equations is generated for the unknown coefficients.

In principle, this set of projections could be performed as is, leaving a system of differential-difference-integral equations for the unknown coefficients $f_l^m(\mathbf{r}, \varepsilon)$. However, the initial substitution of the SH expansion into the various terms of the BTE gives rise to many nonlinear products of SH basis functions. Projecting these nonlinear products by performing the indicated integrations would become unwieldy. Furthermore, since an infinite SH expansion is present, each equation would contain an infinite number of terms, and each equation would therefore be directly coupled to all the other generated equations.

To simplify the system, we take advantage of the SH recurrence relations[3]. These relations allow us to re-express all nonlinear products of SH basis functions in terms of linear combinations. Once each term in the BTE is expressed as a linear combination, we can take advantage of the orthogonality of spherical harmonics and easily perform the projections indicated by Eqn. (2).

IV. Generalized System of SH Equations

After using recursion and performing the indicated integrations, we obtain the remarkable result that almost all of the infinite terms in each equation vanish identically due to orthogonality. Furthermore, we find that the coupling between equations is only through neighbors. Another extremely useful result is that each equation has an identical form. The system can therefore be automatically generated to arbitrarily high order and then be solved numerically. The equation for the l, m SH coefficient is given below. The equation for any of the other SH coefficients is obtained by appropriately changing the value of the indices l, m to other allowed integers:

$$\left\{ \sum_{i=1}^2 v(\varepsilon) \left[\frac{\partial}{\partial x_i} - eE_i \left(\frac{\partial}{\partial \varepsilon} - \frac{l-1}{2} \frac{\gamma'}{\gamma} \right) \right] \hat{a}_i^+ + v(\varepsilon) \left[\frac{\partial}{\partial x_i} - eE_i \left(\frac{\partial}{\partial \varepsilon} + \frac{l+2}{2} \frac{\gamma'}{\gamma} \right) \right] \hat{a}_i^- \right\} f_l^m = \left[\frac{\partial f_l^m}{\partial t} \right]_c \quad (2)$$

where $v(\varepsilon) = \frac{\sqrt{2m\gamma}}{m\gamma'}$, $\gamma' = d\gamma/d\varepsilon$, and γ represents the dispersion relation; the sum is over the 2 directions in the $x - z$ plane; and the operators \hat{a} have been defined as,

$$\hat{a}_1^- f_l^m \equiv \frac{1}{2} \left\{ -\alpha_{l-1}^m \alpha_l^m f_{l-1}^{m-1} + \alpha_{l-1}^{-m} \alpha_l^{-m} f_{l-1}^{m+1} \right\} \quad (3)$$

$$\hat{a}_1^+ f_l^m \equiv \frac{1}{2} \left\{ \alpha_{l+1}^{-m+1} \alpha_l^{-m+1} f_{l+1}^{m-1} - \alpha_{l+1}^{m+1} \alpha_l^{m+1} f_{l+1}^{m+1} \right\} \quad (4)$$

$$\hat{a}_3^- f_l^m \equiv \alpha_{l-1}^{-m+1} \alpha_l^m f_{l-1}^m \quad (5)$$

$$\hat{a}_3^+ f_l^m \equiv \alpha_{l+1}^m \alpha_l^{-m+1} f_{l+1}^m \quad (6)$$

where $\alpha_l^m \equiv \sqrt{\frac{l+m}{2l+1}}$. Due to the need to consider self-consistent boundary conditions in multi-dimensional problems, it is necessary to produce a set of 2nd order equations from this set of 1st order equations. Before doing this, however, it is possible to eliminate mixed partial derivative operators such as $\frac{\partial^2}{\partial \varepsilon \partial x_i}$ that would occur from such a substitution procedure by introducing the gauge transformation,

$$\varepsilon \rightarrow \varepsilon' - e\phi(\vec{x}) \equiv H \quad (7)$$

as discussed in Ref. [2]. Here, the effect of the transformation is to produce the same set of equations as Eqn. (2) except that now, the energy derivatives are no longer present. This is because the derivatives with respect to position now have a new meaning, i.e. they are evaluated for fixed values of H .

V. Results

We have solved this generalized system for the 2-D space-independent case. This system is truncated, discretized and solved numerically. Solving the space independent BTE to 10 orders requires less than a minute of CPU time on a Sun4 work-station. Fig. 1 shows the SH coefficients corresponding to $l = 0, 2, 4$ for a $100 \frac{kV}{cm}$ electric field 45 degrees from the p_3 axis in the 111 direction. Fig. 2 shows a comparison of this result with the case of the electric field in the direction of the p_3 axis. The purpose is to demonstrate how different SH coefficients become important to resolve the 2-dimensional angular dependence of the distribution function. Calculation of the isotropic coefficient remains the same, while the magnitude of higher order coefficients change in response to the changing values of the spherical harmonics. It is believed that some numerical error is present at low energy for the higher order coefficients as it was expected that the shapes of the coefficients of the same order in l should be the same.

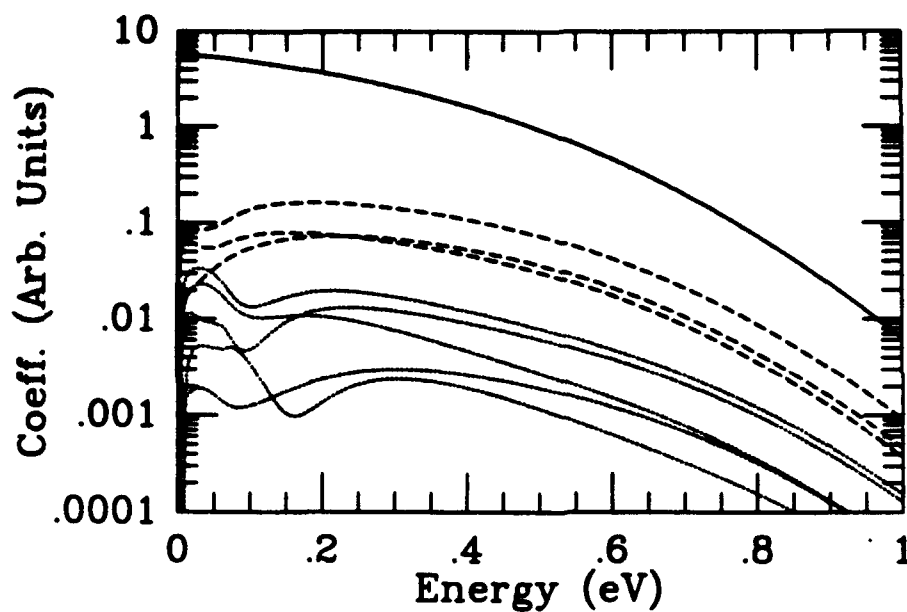


Fig. 1. SH Coefficients corresponding to $l=0,2,4$ for a 100 kV/cm electric field in the 111 direction.

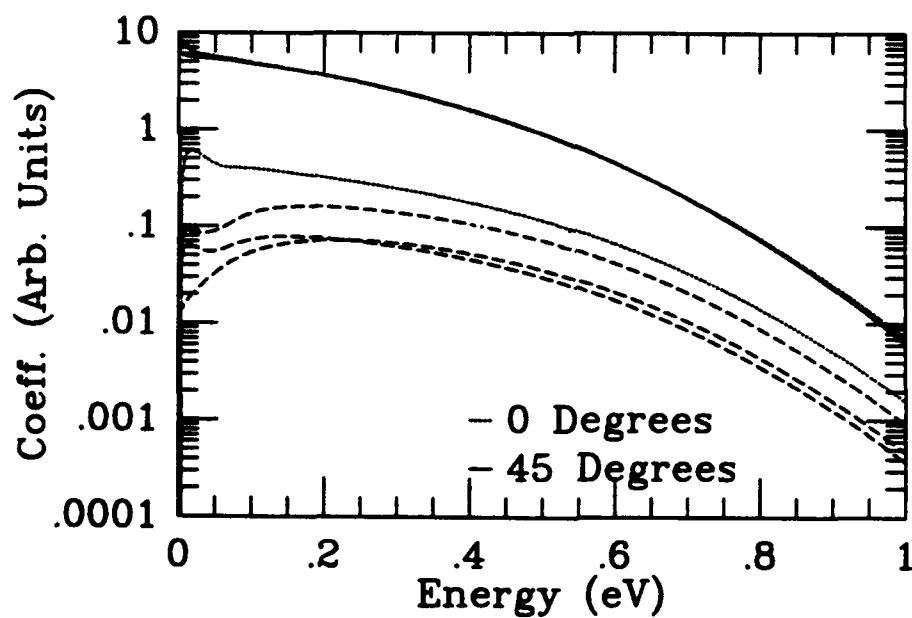


Fig. 2. A comparison of SH coefficients calculated for two different field orientations.

VI. Summary

We have developed a new method for solving the BTE for 2-dimensional geometries. The method reduces the dimensionality of the problem from 5 to 3. Theoretically, the orientation of the coordinate system should not affect the calculation of the distribution function, however in practice with the direct method, a small numerical difference is noticed.

References

- [1] K. A. Hennacy and N. Goldman, "A Generalized Legendre Polynomial/Sparse Matrix Approach for Determining the Distribution Function in Non-Polar Semiconductors." 1993.
- [2] D. Ventura, A. Gnudi, G. Baccarani, and F. Odeh, "Multidimensional Spherical Harmonics Expansion Of Boltzmann Equation For Transport In Semiconductors," *Appl. Math. Lett.*, vol. 5, 1992.
- [3] G. B. Arfken, *Mathematical Methods for Physicists*. New York: Academic Press, 1966.

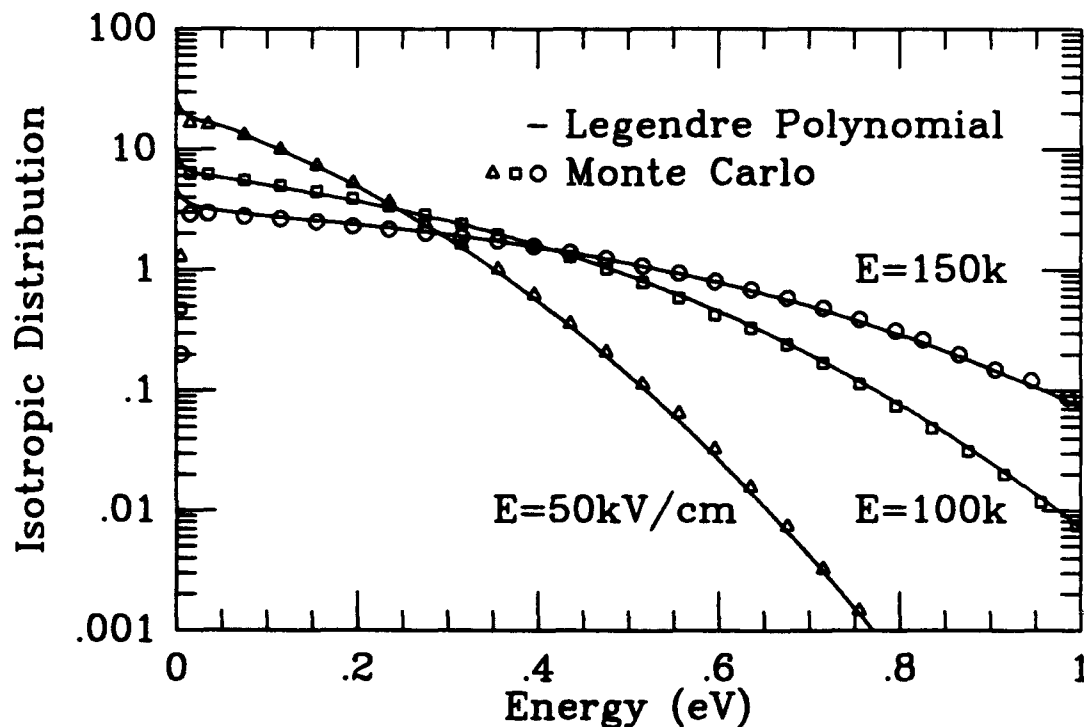


Fig. 3. Isotropic coefficient f_0 for different homogeneous applied electric fields. Comparison of the LP calculation with the Monte Carlo method was excellent up to energies higher than 2eV.

Comparison of Cellular Automata and MINIMOS simulations of submicron MOSFETs

A. Rein, G. Zandler and P. Lugli

Physik Department and Walter Schottky Institut, TU München, D-85747 Garching, FRG

Abstract

We present a detailed comparison of a novel cellular automaton (CA) technique and a standard drift diffusion calculation (MINIMOS) of high field transport in semiconductor devices. The CA method may be viewed as equivalent to the Monte Carlo technique but can easily handle ensembles with more than 10^5 particles, can efficiently deal with complex geometries and achieve accelerations on multiprocessor computers that scale linearly with the number of processors. With this new technique Si MOSFETs have been simulated for different gate lengths and gate voltages and the results compared to MINIMOS.

1. Introduction

Device simulation has become a crucial and strategic part of today's microelectronics [1-3]. Drift diffusion approaches [4] are commonly used for device modelling because of their intrinsic speed. They suffer though from the several approximations of the underlying physical model. The Monte Carlo (MC) method [5,6] belongs on the contrary to the most sophisticated, but at the same time the most costly, of all simulators. For this reason, the MC method still remains restricted to university and laboratory research, and has not yet become a common modeling tool. It would therefore be highly desirable to develop a simulator of comparable physical content as the MC, but much faster and also capable of exploiting the potential offered by vector and parallel processors more naturally.

Recently, a new method has been developed which appears to meet these requisites, the Cellular Automata (CA) approach [7,8]. So far, only preliminary tests of its applicability to device modeling have been carried out. In this paper, a critical assessment of the strengths and limitations of the CA for realistic device simulations is given by presenting the first quantitative and detailed comparison between the CA and a standard Drift-Diffusion approach based on the MINIMOS code [1] for a Si MOSFET.

II. The Cellular automaton for device simulations

Like the MC method, the CA is a physical approach to carrier transport in semiconductors based on the simulation of a random walk of classical particles subject to probability scattering events by phonons, impurities, other carriers, whose energy dependence is evaluated from Fermi's Golden rule [5,6].

In general, the electric field acting on the particles is a function of position; in order to fully account for non-homogeneous situations (and therefore to simulate semiconductor devices) the MC and the CA simulations have to be self-consistently coupled to Poisson's equation [5,6]. Recently, the full BE for carrier transport in semiconductors has been transformed to a CA [7]. This constitutes an important improvement, since CA are traditionally only used in the context of transport, for fluid dynamics [9,10] or drift-diffusion simulations [11]. In general, a cellular automaton consists of a lattice with a finite number of states attached to each lattice

site. The population of these states is simultaneously updated according to deterministic or nondeterministic rules in discrete time steps. The dynamics of CA are governed by local rules, i.e. updating site variables involves only a small number of neighbors in each time step. For this reason, CA can optimally utilize massively parallel computer technology. In addition, the locality of the dynamical rules allows an efficient and flexible treatment of complex geometries. The major characteristics of CA are the two length scales they operate on. The first described by the discrete microworld on a lattice obeying a fictitious dynamics of pseudo-particles, whose length and time scales are much shorter than the physical scales. The second is a continuous macroscale with the physical observables, which are obtained in practice by taking averages over many cells.

In its implementation for the solution of the BE, the CA consists of a lattice in position space, each site of which has a finite number of momentum states. The nondeterministic transition rules between these states associated to collision events are determined from the quantum mechanical scattering rates (in the same way as in MC) and from the classical equations of motion. Due to the locality in position space of quantum mechanical scattering events, which is a basic assumption underlying the BE, there is no principle problem to convert these transitions in momentum space into local CA-rules. On the other hand, the drift and diffusion terms in the BE link the distribution function to its value in different position and momentum space locations, being therefore nonlocal in nature. Additionally, a single semiclassical particle trajectory cannot in general be reproduced exactly on the discretized phase space of the automata. Therefore, the kinetic terms of the BE are replaced by hopping probabilities in both position and momentum space in such a way that the equations of motion are fulfilled on the average for an ensemble of pseudoparticles. By an appropriate choice of lattice constant a and timestep dt , the hopping events can be restricted to transitions between nearest neighbors (or second nearest neighbors). This choice is restricted by the desired resolution in position space and by the maximum physically relevant velocity, which must be less than a/dt . This procedure results in a master equation for the state occupancies which contains only on-site transitions between different momentum states and transitions between nearest neighbors with the same momentum state. For a rigorous derivation and a detailed discussion of the basic algorithm, see [7].

We have estimated the number of operations that translate into MFLOPS required on a scalar processor for simplified MC and CA device simulations on a fixed field distribution that includes nonparabolic bands and the standard scattering mechanisms for GaAs. We find that an ensemble MC simulation requires at least $10^3 \cdot N$ operations, where N is the number of particles in the simulated ensemble, while the CA requires about $10^1 \cdot N$ operations, providing a speed-up of two orders of magnitudes with respect to MC. The principal difference is that the CA maps the Boltzmann equation onto a set of discrete rate equations for discrete variables that involve only integer arithmetic. In momentum space, this discrete dynamics allows one to use predetermined scattering tables both for the total scattering rates as well as for the free flights. In the CA, the final k -state is picked by a simple assignment rather than by the algebraic solution of nonlinear equations as in the MC. In addition, the motion in real space is discrete in the CA and consists of deterministic hops between nearest neighbor cells. Thus, the real space motion is automatically synchronized and occurs in a strictly ordered and predictable fashion, in contrast to the MC, where the exact continuous trajectories of each

particle have to be followed. Clearly, very efficient Poisson solvers need to be coupled to the CA in self-consistent device simulations. Our experience shows that SOR algorithms are too slow and can be the bottleneck of a CA simulation, particularly for large doping gradients. Well known alternatives to SOR methods are also FACR and FFT algorithms [12]. It should be mentioned that the initialization phase of a simulation, on the other hand, requires significantly more effort in the CA method since all scattering tables need to be calculated. Fortunately, it needs to be executed only once. Modern massively parallel computers employing either the message passing or shared memory paradigm offer a substantial reduction in turnaround time, provided efficient algorithms can be implemented on such an architecture. Indeed, both the MC and the CA can be implemented on MIMD (multiple instruction multiple data) machines by assigning subdomains of the whole device to the individual processors [13]. Each processor executes one time step of the simulation on its subdomain, collects the properties of those particles that will leave the subdomain at the end of this time step, passes this information to its neighboring processors, and receives information from them about particles entering its subdomain in the next time step. Such collection of information about the particle exchange with the neighboring processors requires more computations than needed on a single processor. This overhead is the price for the distribution of the computational load onto many processors. We have calculated the ratio of this overhead for the CA and MC method. Let N_{pp} be the number of particles assigned to one processor. Within a MC approach, we estimate that it takes every processor $10^1 \cdot N_{pp}$ operations to collect and digest the information about the outgoing and incoming particles. In the CA, this overhead is independent of the number of particles but only depends on the number of real space cells per subdomain N_{site} and the number of discrete k-states per cell N_k through the relation $10^{-2} N_k \sqrt{N_{site}}$. With typical values for device simulations, $N_{pp} \sim 10^4$, $N_{site} \sim 10^4$, $N_k \sim 10^3$, one finds that the time for the message-related computations required on each processor is two orders of magnitude smaller in the CA than in MC.

III. CA vs. MINIMOS simulation of Si-MOSFET

In order to test the capability of the CA to handle the complex structure of a Si n-MOSFET, a detailed comparison with a standard drift-diffusion algorithm (MINIMOS [1]) has been performed. The CA simulation employs a non-parabolic band structure for electrons and a parabolic dispersion for one effective hole band. Furthermore, optical and acoustic phonon-scattering as well as impact ionization and impurity-scattering are taken into account. Source and drain contacts are $0.24 \mu\text{m}$ long, each separated by 10 nm from the gate contact. An oxide layer of 5 nm thickness has been used (Fig. 1). The doping profile is the one calculated by the MINIMOS pre-processor. The high carrier density in a MOSFET requires large particle ensembles in order to reduce the statistical noise and we have used 300000 particles. On several scalar RISC workstations, the execution time for the CA turned out to be 10–15 times larger than MINIMOS.

In figure 2(a) we show the drain characteristics for two different gate voltages (1.5 V and 2.5 V). The gate length ($L_G = 0.25 \mu\text{m}$) has been chosen close to the limit where MINIMOS simulations can still be expected to be reliable. The drain characteristics agree well with one another, except in the regime of voltages above 2 V , where a slight velocity overshoot (which is not accounted for in the drift-diffusion approach) acts to increase the drain current.

Figure 2(b) shows the I-V characteristics for a shorter gate length of $0.16 \mu\text{m}$. Short

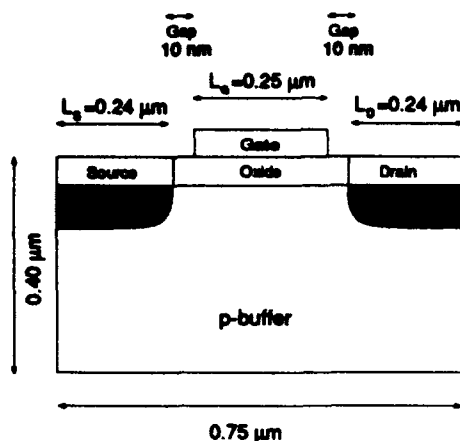


FIG 1. Geometry of a $0.25 \mu\text{m}$ MOSFET. In x-direction the device is subdivided into 126 and in y-direction into 78 blocks, respectively. The n^{++} doping for the contact region is 10^{19}cm^{-3} .

channel effects become more important in this regime, and the two methods give significantly different drain characteristics. This originates almost totally from a significant velocity overshoot that is accounted for by the CA simulation but not by MINIMOS. In fact, as shown in Fig. 3 (a), the electron channel density of the $0.16 \mu\text{m}$ MOSFET is very similar for the two methods up to an applied voltage of 4.0 V and a gate voltage of 1.5 V. In contrast, the drift velocity as obtained in the CA shows a significant overshoot for drain voltages above 0.6 Volt, in contrast to the MINIMOS results (Fig. 3 (b)). At last we present our investigation about impact ionisation. For the cellular automaton we used a microscopic impact ionisation model for high-field energy electron transport [14] and adapt the homogeneous ionisation rate to experimental data [15].

In figure 4(a) and 4(b) position resolved distributions of impact ionisation events are given for the $0.25 \mu\text{m}$ MOSFET at a drain voltage $U_D = 4.0 \text{ V}$ and gate voltage $U_G = 1.5 \text{ V}$ has been used. Figure 4(a) illustrates the MINIMOS simulation, where the maximum of the impact ionisation is located between gate and drain. This also holds for the distribution of impact ionisation using the cellular automaton (Fig. 4(b)), but the maximum is shifted closer to the intersection of gate and drain. The overall agreement is still very satisfactory.

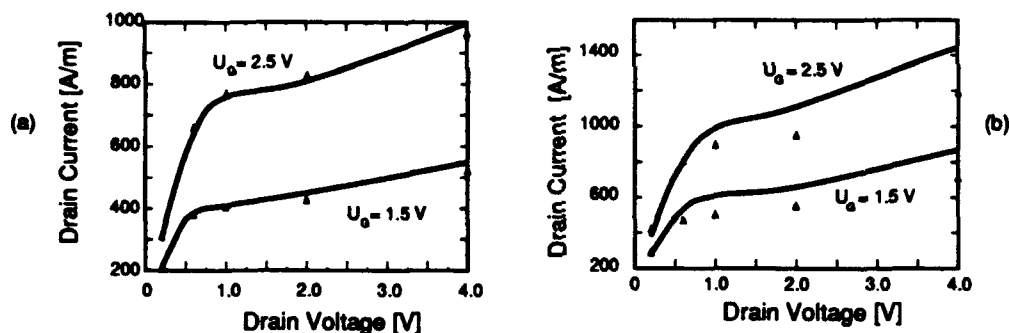


FIG 2. (a) Drain current vs. drain voltage for a $0.25 \mu\text{m}$ Si-MOSFET calculated with the drift diffusion model MINIMOS (triangles) and with the cellular automaton (full line). (b) Drain current vs. drain voltage for a $0.16 \mu\text{m}$ Si-MOSFET.

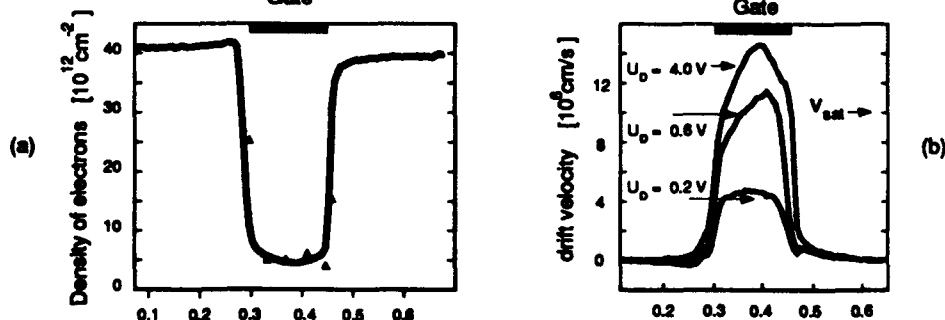


FIG 3. (a) Vertically integrated sheet density of electrons in the $0.16 \mu\text{m}$ MOSFET ($U_D=4.0 \text{ Volt}$; $U_G=1.5 \text{ Volt}$). The full line represents the CA calculation and triangles results of MINIMOS. (b) Vertically weighted drift velocity under the gate (v_{sat} indicates the saturation velocity of bulk silicon).

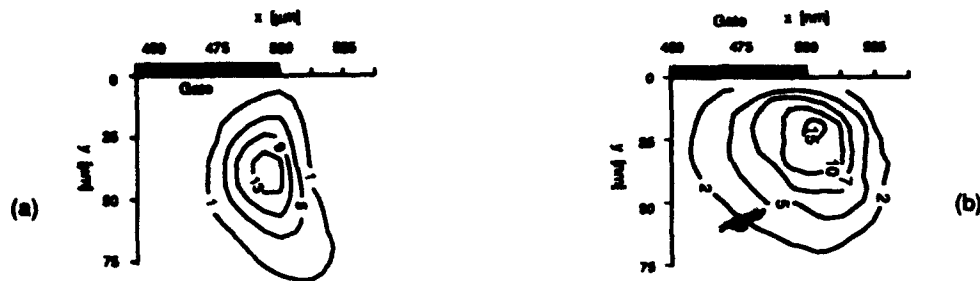


FIG 4. Contour plot of impact ionisation rate per volume [$10^{16} \text{ cm}^{-3} \text{ ps}^{-1}$] for a $0.25 \mu\text{m}$ MOSFET ($U_D=4.0 \text{ V}$, $U_G=1.5 \text{ V}$) in the gate region of the device. The black wide line along the x-axis indicates the gate contact. (a) MINIMOS (b) CA.

IV. Conclusion

We have presented a detailed comparison of a novel cellular automaton (CA) technique and a standard drift diffusion calculation (MINIMOS) of high field transport in semiconductor devices. Good agreement between both methods is found for simulations of a submicron MOSFET within the regime of validity of MINIMOS, while for very short gate length nonlocal transport effects significantly influence the results of the CA simulation. In particular, the occurrence of velocity overshoot in the channel of the device is responsible for the enhanced drain current in the CA simulations.

References

1. S. Selberherr, "Analysis and Simulation of Semiconductor Devices", (Springer, Wien, 1984)
2. W. L. Engl, editor, "Process and Device Modeling" (Elsevier, Amsterdam, 1986)
3. K. Hess, J. P. Leburton, U. Ravaioli, "Computational Electronics" (Kluwer Academic, Norwell, MA, 1991)
4. W. Haensch, "The Drift Diffusion Equations and its Applications in MOSFET Modeling", (Springer, 1991, Vienna)
5. M. Fischetti and S. E. Laux, *Phys. Rev. B* **38**, 9721 (1988)
6. C. Jacoboni and P. Lugli, "The Monte Carlo Method for Semiconductor Device Simulation", (Springer, 1989, Vienna)
7. K. Kometer, G. Zandler and P. Vogl, *Phys. Rev. B* **46**, 1382 (1992)
8. G. Zandler, A. Di Carlo, K. Kometer, P. Lugli, P. Vogl and E. Gornik, *IEEE Electron Dev. Letters* **14**, 77 (1993)
9. U. Frisch, B. Hasslacher, and Y. Pomeau, *Phys. Rev. Lett.* **56**, 1505 (1986)
10. U. Frisch, D. d'Humieres, B. Hasslacher, P. Lallemand, Y. Pomeau, and J. P. Rivet, *Complex Systems* **1**, 649 (1987)
11. M. G. Ancona, *Solid State Electronics* **33**, 1633 (1990)
12. H. Chen, W. H. Matthaeus and L.W. Klein, unpublished
13. R. W. Hockney and J. W. Eastwood, "Computer Simulation Using Particles", (McGraw-Hill, New York, 1981)
14. R. Thoma, H. J. Pfeifer, W.L. Engl, W. Quade, R. Brunetti, and C. Jacoboni, *J. Appl. Phys.* **69**, 2301 (1991)
15. C.A. Lee, R.A. Logan, R.L. Batdorf, J.J. Keimack, W. Wiegmann, *Phys. Rev. A* **134**, 769 (1964)

Two-Dimensional Quantum Modelling of Heterojunction Field Effect Transistors

R. Drury, R.E. Miles, C.M. Snowden

Microwave and Terahertz Technology Group,
Dept. Electrical and Electronic Engineering,
University of Leeds,
Leeds, LS2 9JT, UK

Abstract

A new, fast, two-dimensional model is presented that couples the classical semiconductor transport equations with quantum mechanics. It is particularly suited to the simulation of single channel, pseudomorphic and multi-channel HFETs and delta doped structures, where the primary conduction path between source and drain is via carriers confined to a two-dimensional potential well. The model solves Poisson's and the current-continuity equation self consistently with the effective mass Schrödinger equation, the latter taken in slices perpendicular to the heterojunction. Current flow is modelled by restricting the discretised "quantum" electrons to two-dimensional motion, neglecting non-equilibrium dynamics. The model improves on previously reported versions by allowing a full two-dimensional treatment of the Fermi-level, allowing a non-equilibrium treatment of Schrödinger's equation.

Introduction

The increasing use of AlGaAs/InGaAs/GaAs heterostructure devices has lead to great interest in modelling the electrons confined within 2 Dimensional Electron Gas (2DEG) layers. Classical schemes are based upon the electron wavefunction being described as Bloch states, which in turn are derived from flat band conditions. When events occur that approach the size of the electron wavepacket, ie. the de-Broglie wavelength, this approximation is invalid and quantum effects have to be included. Unfortunately a rigorous solution of the quantum mechanical equations becomes very involved and computationally intensive, limiting their use within physical device modelling. For this reason a simplified scheme has been adopted whereby quantum effects are modelled by solving the effective mass Schrödinger equation (1) across the heterojunction but assuming the classical approximations to hold elsewhere.

$$-\frac{\hbar^2}{2q} \frac{\partial}{\partial y} \left(\frac{1}{m^*} \frac{\partial \xi_k}{\partial y} \right) + (V_{sc} - \lambda_k) \xi_k = 0 \quad (1)$$

where

$$V_{sc} = -\psi + V_h + V_{xc} \quad (2)$$

V_h is the heterojunction potential and V_{xc} the exchange-correlation energy [1]. This approach is justified as the smallest event in the x-direction is the gate which is at least an order of magnitude larger than the electron wavelength. The electron density is now described by equation (3), [2], instead of the classical Fermi-integral.

$$n = N_{c2D} \sum_k |\xi_k|^2 \log_e \left\{ 1 + \exp \frac{q}{kT} (\phi - \lambda_k) \right\} \quad (3)$$

The model therefore solves Poisson's equation self-consistently with the classical charge transport equation, with the modification that the electrons are divided into two classes. The first possessing "quantised motion" are only permitted to move parallel to the heterojunction, and are formed from the electrons whose intersub-band separation is greater than the thermal energy, $k_B T$. The second are normal three-dimensional electrons, which, in order to reduce the computational demands are approximated within the model by the Fermi-integral taken from a quasi-continuous conduction band edge shown in Figure 1.

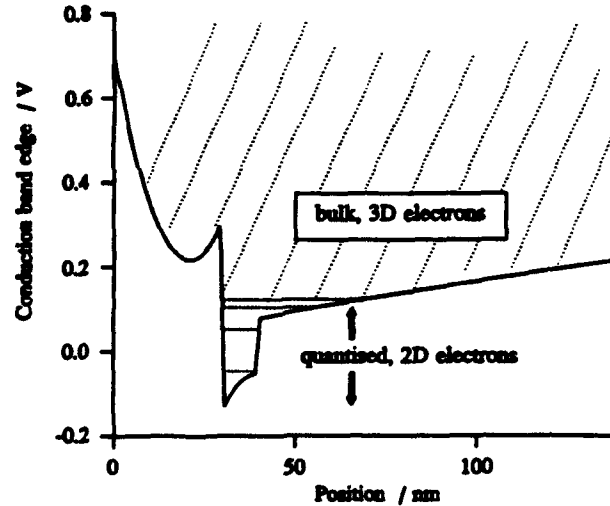


Figure 1. Conduction band edge of a pseudomorphic AlGaAs/InGaAs/GaAs HFET showing the separation of the two- and three-dimensional electrons.

Simulation Details

The equations are discretised over a non-uniform rectangular mesh using central finite differences. Schrödinger's equation is solved, first by finding the eigenvalues via a bisection method based upon a "Sturm" sequence [3]. This method is both fast and robust, with the upper and lower bounds initially set using Gershgorin's theorem [4], and then updated using the last calculated eigenvalue. The eigenvalues are then substituted into Schrödinger's equation and eigenvectors found using a Newton iterative scheme. The current-continuity equation was formulated using current densities calculated at the half-nodes, assuming the independent variables ψ and ϕ vary linearly in between the nodes. This has the advantage that no "Scharfetter-Gummel" interpolation scheme is necessary, substantially simplifying the device equations.

The principle device equations are all highly non-linear in the independent variables ϕ , ψ . Consequently a modified Newton-Raphson iterative scheme was employed which requires the Jacobian, a matrix formed from the partial derivatives of the functions with respect to each of the variables. All of the terms in the two equations, with the exception of the partial derivative

of n_{2D} with respect to ψ are readily differentiable. This term depends explicitly upon the partial derivatives of ξ_k and λ_k , which are calculated using perturbation theory, equation (4)

$$\frac{\partial \lambda_k}{\partial \psi} = -1, \quad \frac{\partial \xi_k}{\partial \psi} = 0 \quad (4)$$

using this result it is evident by inspection of equation (3) that

$$\frac{\partial n_{2D}}{\partial \psi} = \frac{\partial n_{2D}}{\partial \phi} \quad (5)$$

Results

The model has been used to simulate several devices including single channel, pseudomorphic and multichannel FETs. A typical pseudomorphic device structure is shown in Figure 2. The ohmic contacts are assumed to extend to the 2DEG and thus the modelled region is truncated placing the source and drain contacts at the sides. A lumped access resistance is then added explicitly.

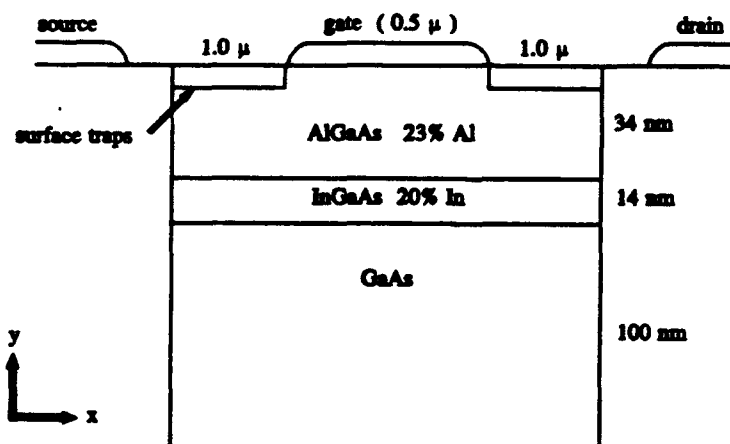


Figure 2. A typical pseudomorphic HFET structure used in the simulations.

Figures 3 and 4 show Fermi-level and conduction band edge profiles, taken at $V_{DS} = 1.7V$ and $V_{GS} = 0V$. The Fermi-level shows significant distortion around the gate, clearly perturbed from equilibrium, although in the InGaAs channel and GaAs substrate the gradient of the Fermi-level (driving force) is nearly parallel to the heterojunction. The conduction band edge shows the two-dimensional potential well formed in the InGaAs.

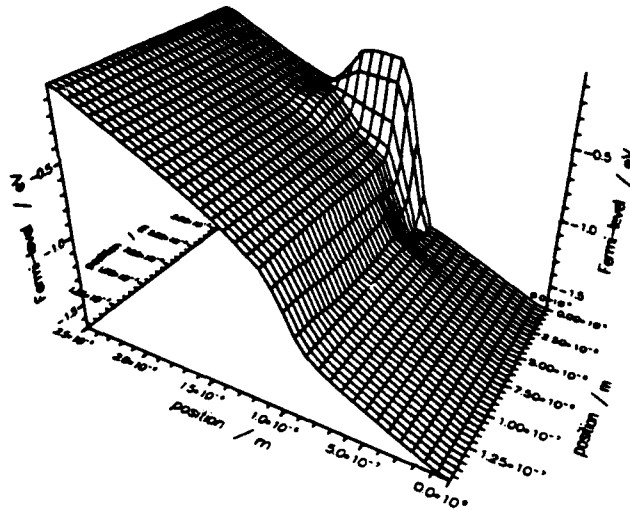


Figure 3. Fermi-level for a pseudomorphic HFET.

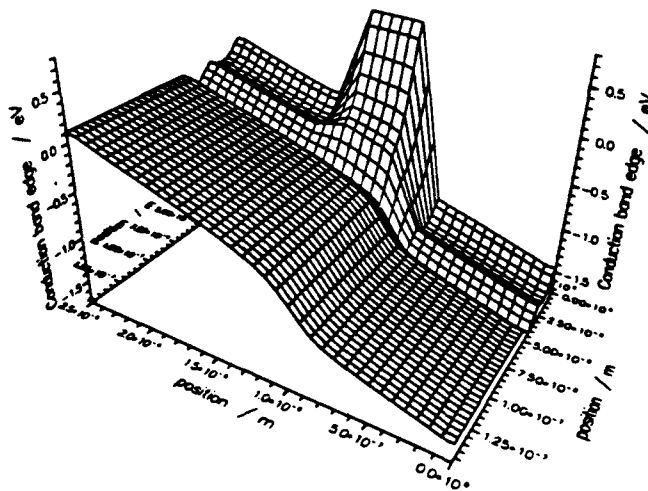


Figure 4. Conduction band edge for a pseudomorphic HFET.

Figure 5 shows the I_{DS} - V_{DS} characteristics for this device, comparing quantum (solid lines) with classical Fermi-integral solutions, from which it is evident that the quantum simulation has a lower output current. This is mainly attributable to the reduced carrier density produced by this scheme, as current flow in the y-direction is mainly limited by the high access resistances to the 2DEG regions which are populated by three-dimensional electrons in both cases.

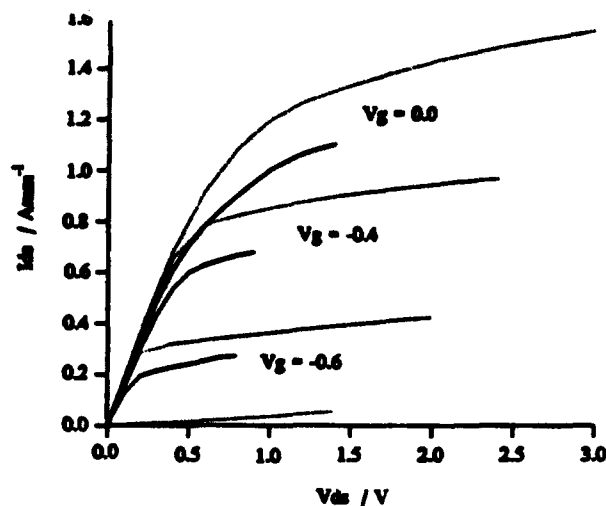


Figure 5. I_{DS} - V_{DS} curves for the quantum and classical simulations.

Conclusions

A two-dimensional HFET model incorporating quantum mechanics is presented that solves Schrödinger's equation in a more self-consistent fashion than has previously been reported. The quantisation in electron motion is explicitly taken into account, and found to have little effect in the final current-voltage curves. A significant reduction in current is observed between quantum and classical models, principally attributable to a lower free electron density.

This work is sponsored by SERC, the Science and Engineering Research Council, UK. and M/A-COM inc., Corporate Research and Development Centre, Lowell, MA 01851-2694, USA. The author also acknowledges Berkeley University USA for the use of 'Sparse', an LU matrix inversion program [5].

References

- [1] F. Stern, "Electron energy levels in GaAs-Ga_{1-x}Al_xAs heterojunction", *Phys. Rev. B* Vol. 30, No.2, pp.840, July 1984
- [2] S. Datta, "Quantum Phenomena", *Addison-Wesley Publishing Co.*, pp. 96, 1989, ISBN 0-201-07956-9
- [3] J.H. Wilkinson, "The Algebraic Eigenvalue Problem", *Springer-Verlag*, pp. 300, 1971
- [4] J. Stöer and R. Burlisch, "Introduction to Numerical Analysis", *Springer-Verlag*, pp. 385, 1980
- [5] K.S. Kundert and A. Sangiovanni-Vincentelli, "A Sparse Linear Equation Solver", Dept. Elec. Eng. and Comp. Sci., University of California, Berkeley, Calif. 94720

Determination of Diffusion Coefficients and Mobilities Using The Effective Field Concept

Ken Hennacy, Neil Goldsman and I.D. Mayergoyz

Department of Electrical Engineering,
University of Maryland,
College Park, MD 20742

Abstract

In recent years, the effective field approach has been developed in terms of hydrodynamic variables [1, 2, 3]. In this paper, the effective field is formulated for the calculation of transport coefficients in a drift-diffusion model. With these transport parameters, it was found that calculations of the energy distribution function tail, and average velocity match suprisingly well with Monte Carlo results, for a variety of cases. Therefore, results so far encourage further investigations into the effectiveness of characterizing device performance along with device reliability in terms of the effective field concept.

I. Introduction

Recently, it has been shown that in some cases the effective field determined from the average energy is sufficient for characterizing the shape of space dependent energy distribution function tails [4]. In this work, the effective field is defined as the table of electric field values that are generated by finding, at each space point, the best match between a homogeneous field and inhomogeneous field calculation of the energy distribution function tail. Then, the effective field concept is extended to include computations of mobilities and diffusion coefficients, with justification based upon the Bogoliubov ansatz [5].

II. Modeling

In 1-D, the following analytical formula for the effective field has been obtained from Boltzmann's equation,

$$E_{eff}(x) = \int_{x_0}^x dx' \frac{dE}{dx'} \left(1 - e^{-\frac{x-x'}{\lambda}} \right). \quad (1)$$

Here, it is seen that the derivative of the electric field is being attenuated by an exponential term, thus demonstrating that the faster the field varies, the more the effective field will lag behind the locally applied field. This behavior is in qualitative agreement with the observations of Ref. [6]. The quantity λ is referred to as the relaxation length, and is shown to be well characterized by the ratio of the diffusion coefficient to the average velocity \bar{u} ,

$$\lambda = \left| \frac{D}{\bar{u}} \right|. \quad (2)$$

This is a transcendental equation, however approximations can be introduced by considering Fig. 1, which is a plot of the ratio of diffusion coefficient to drift velocity for a homogeneous applied field. This plot characterizes the variation of λ , and it is seen that at high fields, where velocity saturation is present, λ varies slowly.

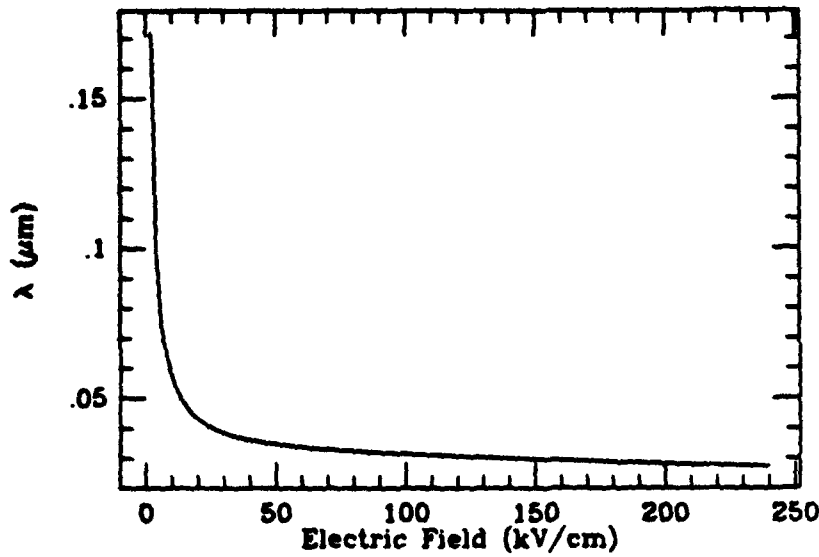


Fig. 1. This graph demonstrates the general dependence of λ upon the size of the electric field. For high fields, λ becomes the length scale over which the average energy relaxes.

Once the effective field is known, the drift velocity is found to have the following simple form,

$$\bar{v}_d = \frac{\bar{E}}{E_{eff}} v_h(E_{eff}) \quad (3)$$

which is a formula that is not restricted to 1-D considerations. Here, $v_h(E_{eff})$ represents the average velocity of carriers in a homogeneous field E_{eff} . For low fields, where the slope of $v_h(E_{eff})$ is fairly constant, this expression reduces to the conventional, local expression for carrier mobility. In the limit that the applied field no longer varies, this expression yields the correct steady state velocity.

Though the drift velocity demonstrates the presence of velocity overshoot, and its relationship to a varying field, there are diffusive and convective contributions to the average velocity as well. In 1-D, the following drift-diffusion-convection (DDC) equation first introduced by Sanchez [6] was used to resolve these components,

$$\frac{d}{dx}(Dn) + nv_d + n \left\langle v_{||}^2 \frac{\partial \tau}{\partial x} \right\rangle = \frac{J_o}{q}. \quad (4)$$

The presence of the $\frac{\partial \tau}{\partial x}$ term is due to the effect of a spatial variation in the collision

rate caused by variations in the donor impurities. Using the effective field concept, the following expressions for the transport parameters were obtained:

$$D(x) = - \int d\vec{p} v_{\parallel}^2 \tau(x, \epsilon) f_s^h(x, \vec{p})$$

$$\langle v_{\parallel}^2 \frac{\partial \tau}{\partial x} \rangle = \int d\vec{p} v_{\parallel}^2 \frac{\partial \tau(x, \epsilon)}{\partial x} f_s^h(x, \vec{p})$$

where f_s^h is the distribution function obtained from the effective field. Although equation (4) represents an initial value problem, it did not pose any major problems for self-consistent device simulations where two boundaries are present. A second order version of this equation was also implemented.

III. Results

Shown in Fig. 2 is the effective field profile generated from a Monte Carlo analysis of the channel region of a MOSFET [3]. Also plotted is the effective field generated from formula (1), exhibiting a remarkable agreement. Shown in Fig. 3 is the effective field profile determined from a case studied in Ref. [7]. Also plotted is the expression of $v_h(E)$ and v_d to demonstrate that there is a significant difference between using the effective field drift term, and assuming that the drift velocity depends only upon the locally applied field.

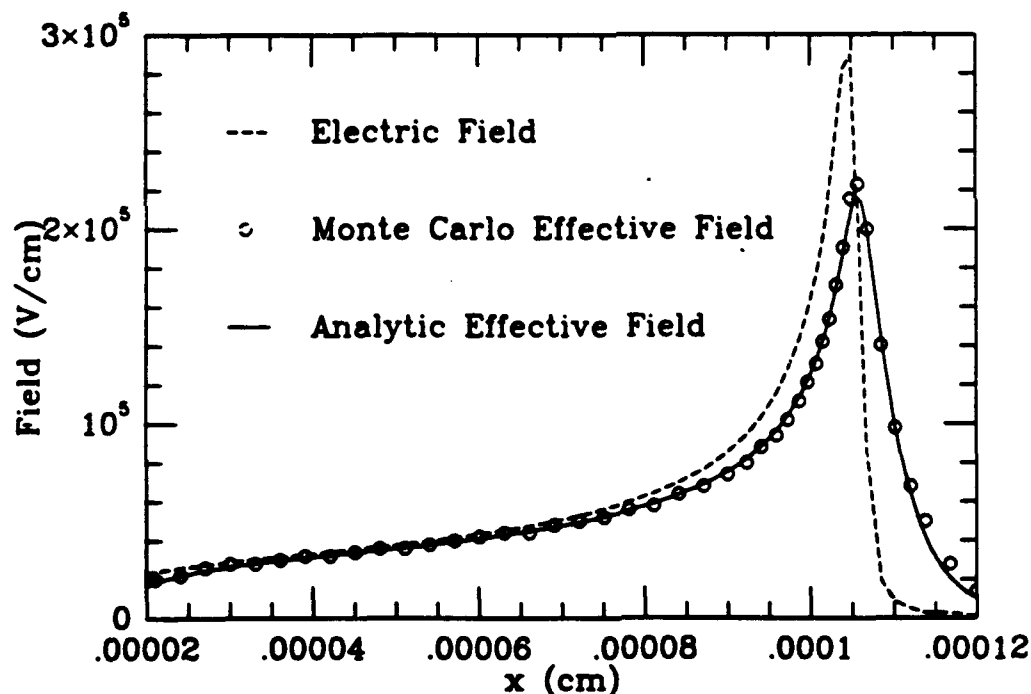


Fig. 2. A comparison of effective fields from ref. [1] and formula (1). A Monte Carlo simulation was used to verify the value of the effective field for most of the points in ref. [1].

In Fig. 4, a comparison of the average velocity obtained from equation (4) and a Monte Carlo calculation is performed. For additional comparison, the first order (P_1) Legendre polynomial calculation performed in Ref. [7] is also included. As can be seen, good agreement between the Monte Carlo calculation and the effective field calculation of the average velocity is obtained. The agreement is better than the P_1 calculation probably due to an implicit avoidance of truncation error in the effective field approach.

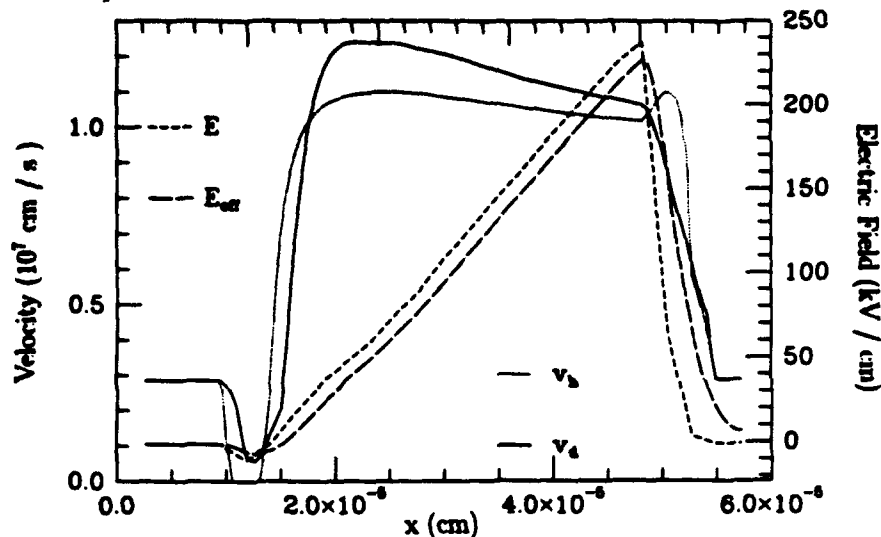


Fig. 3. The drift velocity (v_d), homogeneous field velocity (v_h), self-consistent electric field (E), and effective field (E_{eff}) are plotted for a n^+nn^+ diode simulation, Bias = 5.0 Volts. The discrepancy between v_h and v_d demonstrate the importance of the non-local dependence of the mobility.

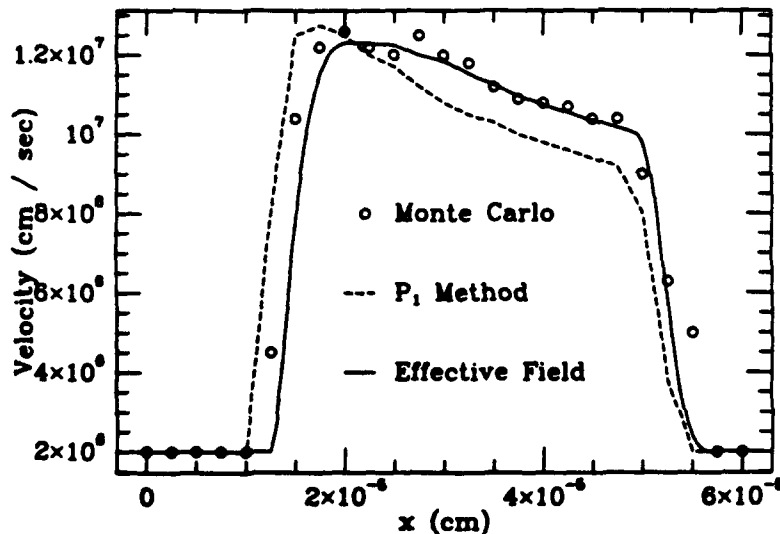


Fig. 4. The average velocity from a drift-diffusion (DD) equation, a Monte Carlo (MC) simulation, and a 1st order Legendre polynomial (P_1) calculation found in ref. [4] are plotted for comparison. Better agreement than the P_1 approach is attributed to an implicit avoidance of truncation error.

The discrepancy in the barrier region of the device is due to the fact that the diffusion coefficient has been overestimated by the effective field formulation presented so far. To account for temperature reduction due to the barrier, the effective field is reformulated using a different approximation scheme. This approximation scheme insures that the correct equilibrium behavior is obtained, and predicts temperature lowering. However, a detailed comparison and study in this region of the device is left for future work.

III. Summary

An approach for device modeling using the drift-diffusion model has been developed which does not use any empirically determined parameters except for those that are associated with deformation potentials and band structures. Considerations so far have been restricted to a single-valley model of silicon. However, the generality of the approach encourages an investigation into considerations of multiple bands and anisotropic effects.

References

- [1] N. Goldsman and J. Frey, "Electron Eenergy Distribution for Calculation of Gate Leakage Current in MOSFETs," *Solid-State Electronics*, vol. 31, no. 6, pp. 1089-1092, 1988.
- [2] C. Fiegna, F. Venturi, E. Sangiorgi, and B. Ricco, "Simple and Efficient Modeling of EPROM writing," *IEEE Transaction on Electron Devices*, vol. ED-38, pp. 603-610, 1991.
- [3] E. E. Kunhardt, J. Wu, and B. Penetrante, "Nonequilibrium macroscopic descriptions of electrons in weakly ionized gases," *Physical Review A*, vol. 37, p. 1654, 1988.
- [4] S.-L. Wang, N. Goldsman, and K. Hennacy, "Determination of Space-Dependent Electron Distribution Function by Combined Use of Energy and Boltzmann Transport Equations: Improvement, Evaluation, and Explanation," *IEEE Transactions on Electron Devices*, vol. 37, p. 1821, 1992.
- [5] E. E. Kunhardt, "Macrokinetic theory of electrons in gases," *Physical Review A*, vol. 43, p. 6664, 1991.
- [6] T.-W. Tang and S.-C. Lee, "Transport Coefficients For A Silicon Hydrodynamic Model Extracted From Inhomogeneous Monte-Carlo Calculations," *Solid-State Electronics*, vol. 35, p. 561, 1992.
- [7] M. Sanchez, "Carrier Heating Or Cooling In Semiconductor Devices," *Solid-State Electronics*, vol. 16, p. 549, 1973.
- [8] H. Lin, N. Goldsman, and I. D. Mayergoyz, "Device Modeling by Deterministic Self-Consistent Solution of Poisson and Boltzmann Transport Equations," *Solid-State Electronics*, vol. 35, no. 6, pp. 769-778, 1992.

ALGORITHMS FOR THE SOLUTION TO THE HELMHOLTZ EQUATION IN THE NUMERICAL SIMULATION OF SEMICONDUCTOR LASERS.

M. Gault P.A. Mawby and M.S. Towers
Department of Electrical and Electronic Engineering
University of Wales
Swansea, UK.

Abstract

In the numerical simulation of semiconductor lasers a fast accurate method is essential for the solution to the Helmholtz equation. Various algorithms are considered for this purpose in terms of speed, efficiency and accuracy and a new algorithm is developed based on the effective index method with a novel solution to the algebraic eigenvalue equation. The algorithm is found to be highly efficient for all modes.

I. Introduction

As laser structures increase in complexity it becomes more important to simulate their behaviour in order to understand phenomena dependent on structure and hence optimise their performance. As manufacturing techniques improve the buried heterostructure (BH) device is likely to grow in importance. A typical BH device is shown schematically in figure 1 in which the InGaAsP active region is surrounded by the wider band gap InP on all sides. In the simulation of these structures the wave equation must be solved to obtain the optical intensity of each lasing mode and the aim of this paper is to investigate the most efficient way to solve this equation.

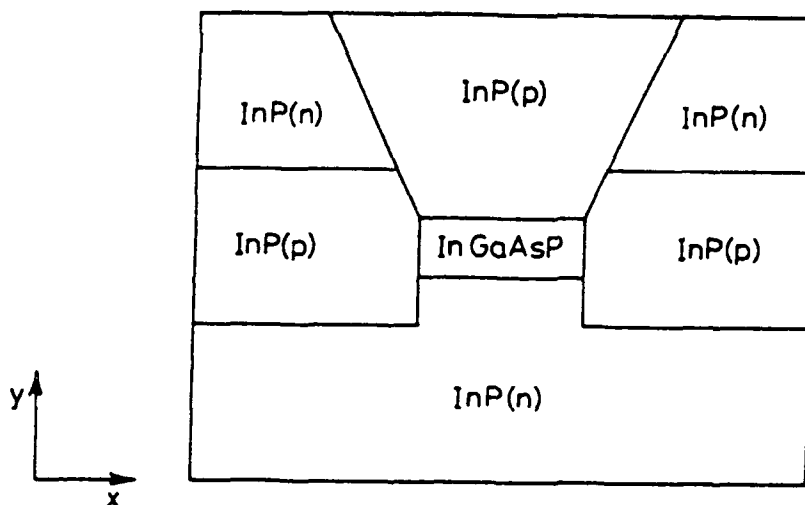


Figure 1 Device Structure

II. Modelling

Assuming TE modes dominate then the wave equation may be given in scalar form as

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \epsilon(x,y)k_0^2 - \beta^2 \right) E_x = 0 \quad (1)$$

where $\epsilon(x,y)$ is the permittivity (which may be complex due to gain or loss), E_x is the optical field, k_0 is the free space wave number and β is the propagation constant. This equation must be solved in two

dimensions for the structure in figure 1. Of the efficient methods that exist the effective index method [1] is the most commonly used, however as the active area decreases in size and the solution approaches cut-off the accuracy deteriorates [2]. The weighted index method [3] is an alternative approach which attempts to increase the accuracy of the solution whilst avoiding the complexity of a two-dimensional method.

The Weighted Index Method

The weighted index method is an improvement on the effective index method as it uses a weighted mean of the permittivities in the y direction for each node in the x axis and a weighted mean of the permittivities in the x direction for each node in the y axis. This is described with reference to figure 2.

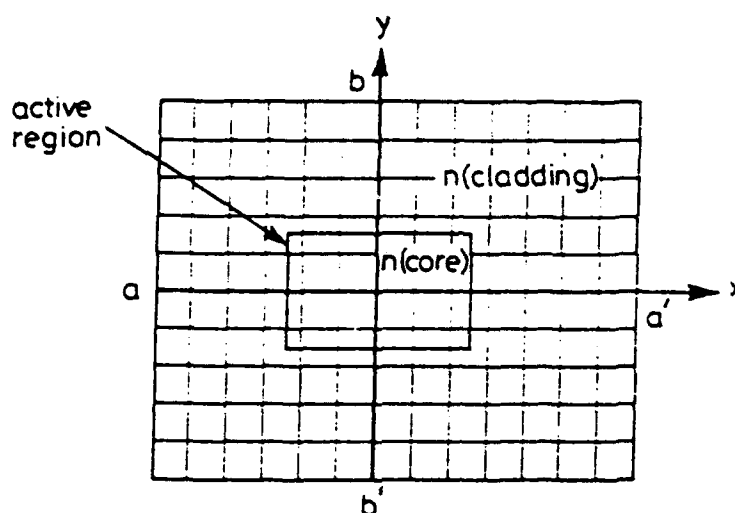


Figure 2 The Weighted Index Method

The best trial solution is found for

$$E = F(x)G(y) \quad (2)$$

F and G are complex functions of the single variables x and y respectively and satisfy the ordinary differential equations

$$\frac{d^2 F}{dx^2} + K_F F = \beta_x^2 F \quad (3)$$

$$\frac{d^2 G}{dy^2} + K_G G = \beta_y^2 G \quad (4)$$

where β_x and β_y are the propagation constants of the differential equations. Equation 3 represents a cut along aa' and equation 4 represents a cut along bb' . There are p nodes along the aa' and q nodes along bb' . K_F is the weighted mean of $\epsilon(x, y)k_0^2$ for each nodal value x_p . The mean is taken along the y direction using weights w_{yq} , i.e.

$$K_p = k_0^2 \sum_q w_q e_{pq} \quad (5)$$

In a similar fashion K_G is given by

$$K_G = k_0^2 \sum_p w_p e_{pq} \quad (6)$$

The weights are defined optimally as

$$w_p = \int_{x_1}^{x_2} F^2 \cdot dx \quad (7)$$

$$w_q = \int_{y_1}^{y_2} G^2 \cdot dy \quad (8)$$

where F and G have undergone the normalisation

$$\int F^2 = 1 \quad (9)$$

$$\int G^2 = 1 \quad (10)$$

The best value of β^2 is given by the Rayleigh quotient [3]

$$\beta^2 = \beta_x^2 + \beta_y^2 - k_0^2 \sum_p \sum_q w_p w_q e_{pq}^2 \quad (11)$$

The two one-dimensional wave equations (3,4) are coupled via the weighting factors and are solved alternately until the value for β converges.

The algebraic eigenvalue equation

With both the effective and weighted index methods a one dimensional algebraic eigenvalue equation must be solved.

$$AE = \beta^2 E \quad (12)$$

where A is complex and tridiagonal and the optical field E is a column vector. If a non-uniform grid is used then A is unsymmetrical and must be symmetrised using a similarity transformation. The eigenvalue and eigenfunction of the coefficient matrix A give the propagation constant β and the corresponding optical field E respectively. A new highly efficient method is described next.

Evaluation of the characteristic polynomial

A very fast method to obtain any required eigenvalue is the evaluation of the characteristic polynomial. A trial value of the eigenvalue is used in a Sturm sequence which can then be used to locate the required eigenvalue via bisection. For any trial eigenvalue λ , the Sturm sequence is defined as :-

$$\begin{aligned} p_0(\lambda) &= 1 \\ p_1(\lambda) &= (a_1 - \lambda)p_0(\lambda) - a_2^2 p_{-1}(\lambda) \end{aligned} \quad (13)$$

The number of changes in sign between consecutive numbers of the sequence is equal to the number of eigenvalues smaller than λ . The advantage of this method is its flexibility in that it allows all the eigenvalues to be found in a given range however the disadvantage is it cannot cope with complex matrices.

An alternative method is inverse power iteration with successive eigenvalue refinement. It consists of iterating around the equation

$$(A - \mu_i I)x_{i+1} = k_i x_i \quad (14)$$

where k_i is chosen such that $\|x_{i+1}\|_\infty = 1$. μ_i is adjusted on each iteration via

$$\mu_{i+1} = \mu_i + \frac{1}{k_i} \quad (15)$$

A linear system of equations must be solved at each step. This algorithm is fast and can cope with complex eigenvalues and eigenvectors. The main disadvantage is that the method requires good approximations to both the eigenvalue and the eigenvector of the problem to initialise. However advantage can be taken of the fact that the change in imaginary components of the permittivity is much less than the change in real components. This is due to the BH laser being index-guided in both lateral and transverse directions. To obtain a good initial approximation to the complex problem the imaginary components can be ignored and the real problem can be solved via the evaluation of the characteristic polynomial for the eigenvalue and an inversion performed for the eigenvector. The results can then be used to initialise the complex inverse iteration with successive eigenvalue refinement algorithm to obtain the complex solution. A flowchart of the algorithm is shown in figure 3

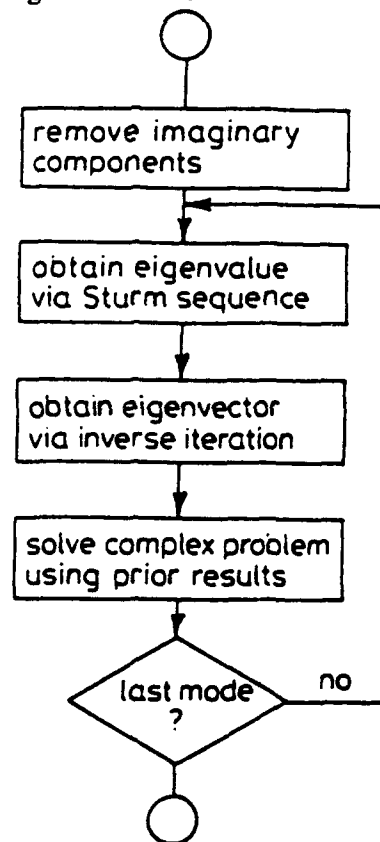


Figure 3 Flow Chart of Operation

III. Results and Discussion

Results are shown in figure 4 for an InGaAsP/InP device using the effective index Method (EI), the Weighted Index Method (WI) and a two-dimensional solution (2D)[2]. The fundamental and first order mode indices are calculated for varying channel widths (W). In this example where the core is 0.35 microns thick the effective index method becomes significantly in error when the width is less than two microns i.e. when the active region area is less than approximately $0.6 \mu\text{m}^2$.

The significance of the above results will now be considered with respect to the design of BH lasers. One of the most important characteristics of these devices is the power at which the first kink in the light-current characteristics occurs. This kink is due to the first order mode achieving gain and commencing lasing. The above results indicate that for accurate determination of this lasing power the EI method is insufficient and the WI method is preferred. An interesting implication of the results given in figure 4 is that the EI method will always underestimate the optical power at which the first order mode starts to lase and the weighted index method will always overestimate this power. This is a direct result of the effective index method overestimating the mode index and hence overestimating the optical gain. Equivalently, the weighted index underestimates the mode index and thereby underestimates the optical gain. This has important implications for device engineers. If a laser is designed to have a minimum output power at which the first order mode starts to lase, then using both the WI and EI methods, they can be reasonably sure that the experimental power will be bounded by the results calculated from the two methods.

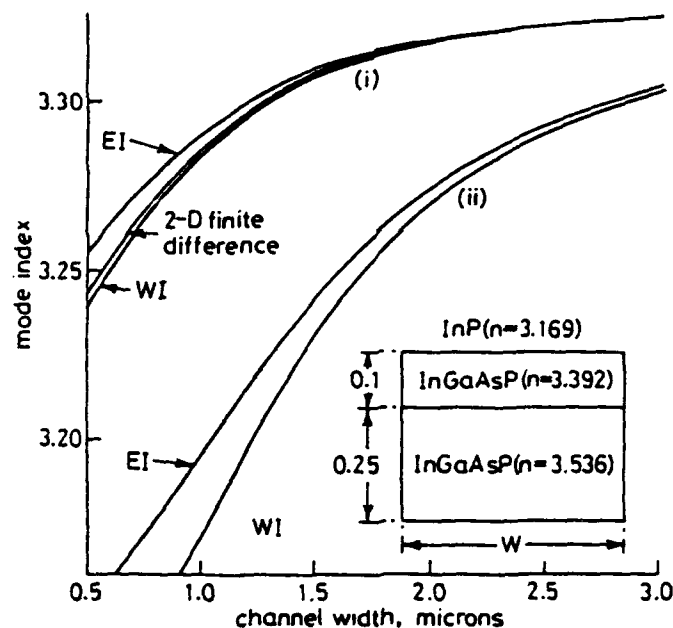


Figure 4 Typical Result

(i) Fundamental mode
(ii) First-order mode
(INSET: Device structure)

References

- [1] BUUS, J.: 'The effective index method and its application to semiconductor lasers', *IEEE J. of Quantum Electronics.*, QE-18, pp 1083-1089, 1982.
- [2] SEKI, S., YAMANAKA, K., and YOKOHAMA K.: 'Two dimensional analysis of optical waveguides with a non-uniform finite difference method', *IEE Proc. J, Optoelectron.*, 138, (2), pp. 123-127, 1991.
- [3] KENDALL, P., ADAMS, M.J., RITCHIE, S. and ROBERTSON, M.J.: 'Theory for calculating approximate values for the propagation constants of an optical rib waveguide by weighting the refractive indices', *IEE Proc. Pt. A*, VOL. 134, (8), pp 699-702, 1987.

Visualisation Techniques for Semiconductor Bandstructures

Alan Beck, Franklin Bodine, P. D. Yoder, and Umberto Ravaioli

Beckman Institute
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA

Abstract

Visualisation of semiconductor bandstructures is accomplished in the 3-D Brillouin zone by using color rendering techniques to generate energy isosurfaces and color contour plots on crosssectional planes. Combination of the various techniques offers an efficient way to analyze the data and get physical insight relevant for carrier transport. The visualization is performed on a Digital 5000 workstation using the Application Visualisation System (AVS) graphics software.

I. Introduction

Advanced investigations of transport phenomena in semiconductors require a detailed knowledge of the energy bandstructure. This is particularly true when high energy effects are investigated, since available analytical formulations of the bandstructure are inaccurate or incomplete. For instance, some Monte Carlo applications suitable for the simulation of impact ionization and of other effects relevant for device reliability, implement algorithms with a full bandstructure for the evaluation of momentum space trajectories. We found that 3-D visualisation of the bandstructure in momentum space is an important step for the understanding of the transport physics and for the development of optimization techniques for this type of Monte Carlo simulations. Since scattering rates are also closely related to the bandstructure, the same visualization techniques can be applied to analysis of the momentum dependence of scatterings.

We have experimented with a number of available visualisation tools and we have found the Application Visualisation System (AVS) software to be very suitable for these 3-D visualization problems. In the AVS environment, most of the applications are built by an interactive symbolic procedure, assembling an application network with modules available in a menu. Custom applications can be developed by writing additional modules using C language programming. An example of AVS application network we have used to visualize the bandstructure of silicon is shown in Fig. 1.

II. Visualization Approach

The approach which we found most effective to visualize energy bandstructures is an animation of energy isosurfaces. We have built a Brillouin zone and we display the isosurfaces in the cube which contains it. The slight redundancy of information which falls outside the zone helps in getting a visual appreciation of the connection between adjacent Brillouin zones. Since an isosurface only displays information for one energy value at a time, it is

often necessary to combine this technique with others to get a more complete idea of the overall behavior. Several isosurfaces can be displayed at one time with the use of transparency and color coding, although this may lead to extremely complicated images. This type of visualization is better suited to visualization of only the irreducible wedge of the Brillouin zone. On a full scale, we found very useful to combine energy isosurfaces with cross-sectional planes where a color contour plot of energy is represented. The isosurfaces are generated only on one side of the plane, which can be made to slide back and forth, and the intersections between the surfaces and the plane offer very helpful visual cues. The capability to perform rotations of the solid figure allows then the viewer to pinpoint any specific region of interest. An example of AVS application network we have used to visualize the bandstructure of silicon is shown in Fig. 1. The network represents actual subroutines which are assembled in a symbolic way. The bandstructure data is input through the branch beginning with the block *read field* while the Brillouin zone is constructed by the adjacent branch beginning with *read geom*. The image is displayed after the program runs through all the blocks, and the user can then interactively modify the image attributes, like orientation, size, color palette, lighting, to adjust the image, or can run again the program by simply modifying visualization parameters, like the value of energy isosurface, the position of the cross-sectional plane, or the data displayed in the given geometry (e.g. switch from bandstructure to scattering rates). The program can also be changed interactively, by breaking the connections in the network, and adding or subtracting modules. A large amount of information can be visualized at one time, since the isosurfaces can be color coded with another parameter (e.g. scattering rate) so that by scanning isosurfaces at different energies a 5-D space is represented.

We present here several examples of 3-D visualization for the silicon energy bandstructure. Figure 2a shows a view of a cross-sectional 100 plane for the first branch of the conduction band in the full Brillouin zone. For the sake of reproduction, we use here a gray-scale palette, ranging from black (lowest energy) to light gray (highest energy). For color representation, we normally use the conventional rainbow coloring scheme, ranging from blue to red. It is easy to detect the four X-valleys situated on this plane in correspondence of the darkest regions. The same image for the second branch of the conduction band is shown in Fig. 4a. In both images, the energy isosurfaces are behind the cross-section. On the computer, the position of the plane can be moved interactively in real time by translation or rotation. Figures 2b, 3a,b and show the full isosurface configuration, at various energies, for the first conduction band branch, and Fig. 4b for the second branch. Figure 5 and 6 show results for the conduction band of gallium arsenide. The silicon-like X-valley isosurfaces can be easily identified by comparison with the previous images.

The data for the bandstructure were obtained by using an empirical pseudopotential approach. All the images are based on a cubic uniform mesh with 41 nodes on each side. The visualization was performed with AVS software on a Digital 5000 workstation and the images have produced on a Tektronix dye sublimation printer.

Acknowledgments - This work has been partially supported by the Semiconductor Research Corporation and by the National Science Foundation (NSF ECS 91-22768).

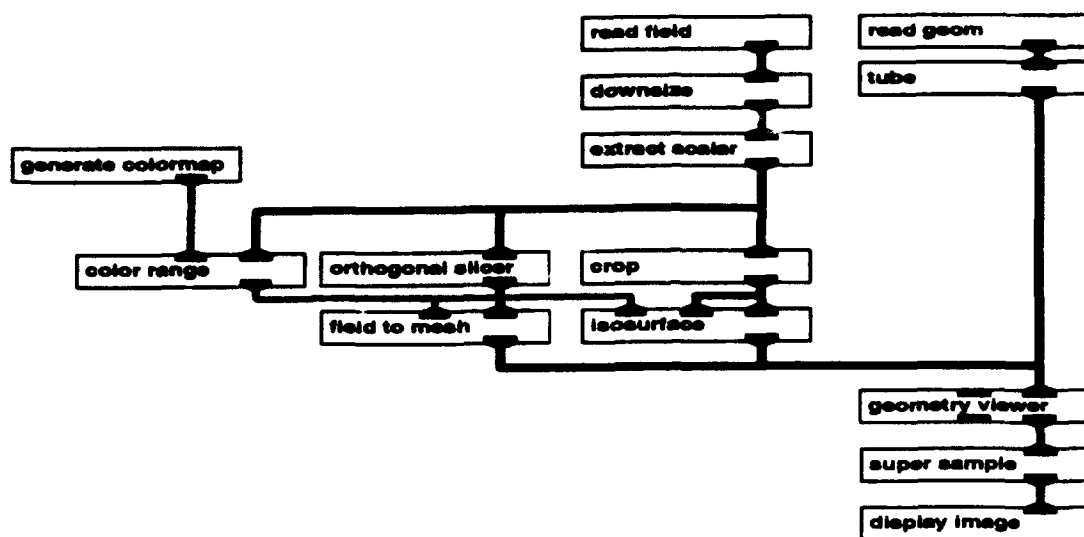


Figure 1. Simplified AVS network used for the visualization.



Figure 2. (a) Gray-scale contour plot of the first conduction band for silicon on the {100} plane; (b) Energy isosurfaces at $E = 0.2$ eV.

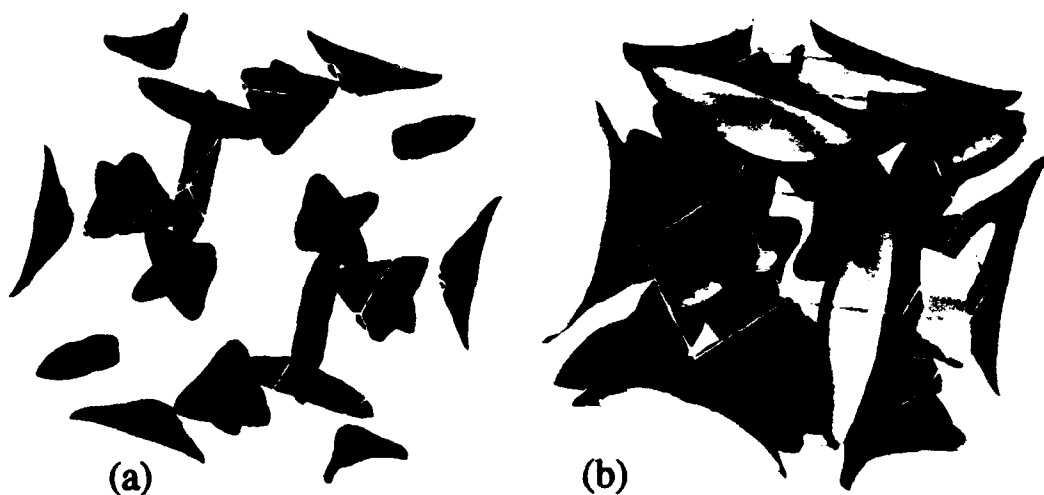


Figure 3. Energy isosurfaces of the first conduction band for silicon at (a) $E = 0.5 \text{ eV}$ and (b) $E = 1.5 \text{ eV}$.



Figure 4. (a) Gray-scale contour plot of the second conduction band for silicon on the $\{100\}$ plane; (b) Energy isosurfaces at $E = 2.4 \text{ eV}$.

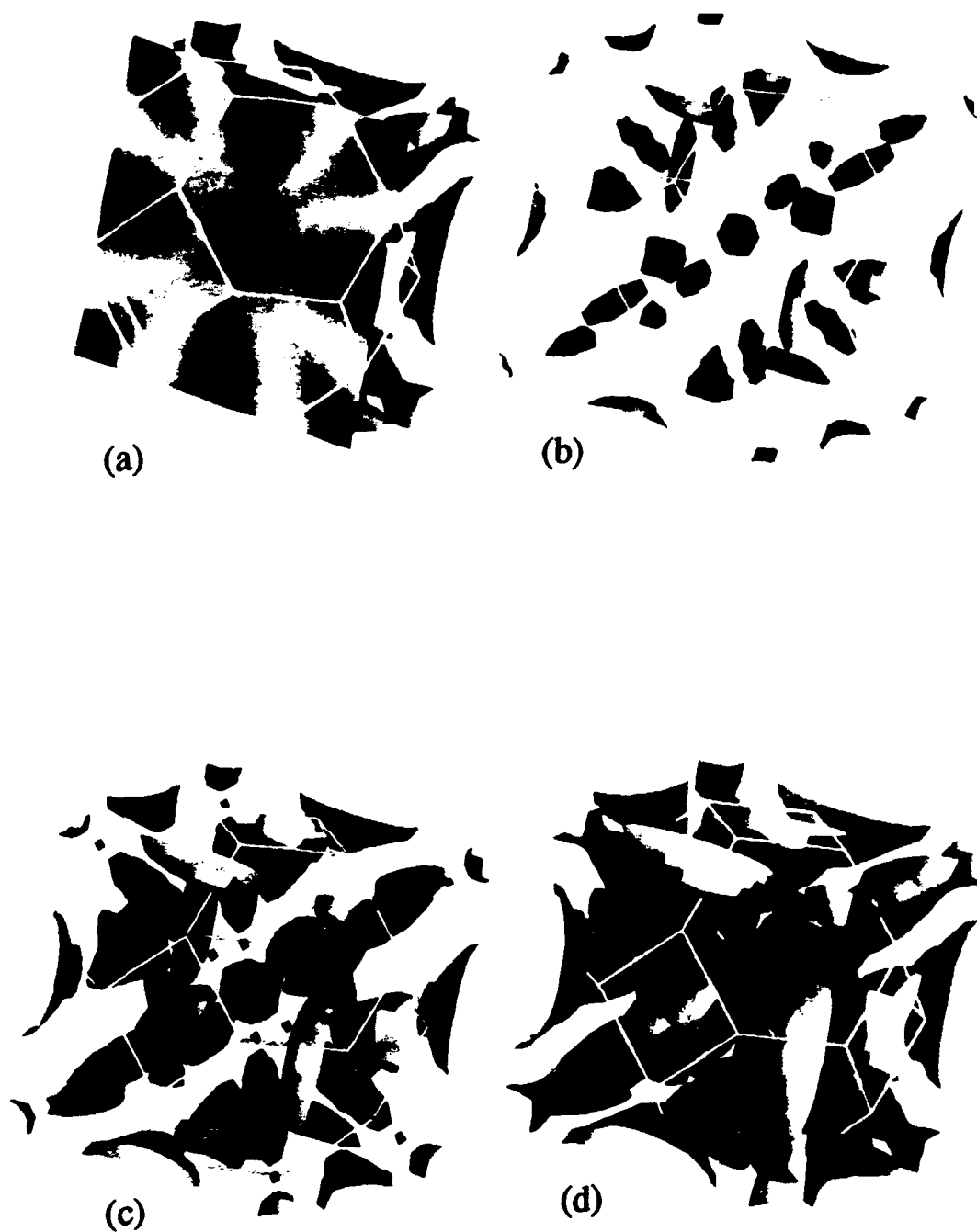


Figure 5. (a) Gray-scale contour plot of the conduction band for GaAs on the $\{100\}$ plane; Energy isosurfaces at (b) $E = 0.75 \text{ eV}$; (c) $E = 1.1 \text{ eV}$; (d) $E = 1.5 \text{ eV}$.

Computation of the Electron Distribution Function Employing the Theory of Stochastic Differential Equations

Can E. Korman, Alfredo Piazza and Pattana Rugkwamsook

Department of Electrical Engineering and Computer Science
The George Washington University
Washington, DC 20052

Isaak D. Mayergoyz

Electrical Engineering Department and Institute for Advanced Computer Studies
University of Maryland
College Park, Maryland 20742

Abstract

In a previous paper, it has been suggested to employ the theory of stochastic differential equations (SDE) for the modeling of electron transport in semiconductors. It was shown that the differential equations which describe electron transport in semiconductors, and which are used in Monte Carlo simulations, can be interpreted as stochastic differential equations driven by inhomogeneous randomly weighted Poisson processes. Based on this connection, the theory of SDE was employed to demonstrate that the forward Kolmogorov-Feller equation which characterizes the transition probability density function of this random process can be integrated over the initial conditions to obtain the linear Boltzmann transport equation.

In the paper, we will expand on our previous results and present a transport model which leads to the efficient numerical computation of the electron distribution function. The model is based on an approximation of the collision integral in the Boltzmann transport equation by differential operators. As a result of this approximation, the electron distribution is characterized by a second order partial differential equation in momentum space. It turns out that the coefficients of this partial differential equation are the first and second order moments of the scattering transition rate. We will present the derivation of these terms and also present some preliminary results on the computation of the electron distribution function based on this model. Furthermore, by comparing these results with Monte Carlo simulations, we will explore the limits of applicability of this model.

I. Background:

Previously, it has been suggested that the theory of stochastic differential equations can be employed to model current transport in semiconductors, (see Ref. [1]). In this section, we will review some of the previous results and establish a connection between

semi-classical transport theory and the theory of stochastic differential equations. According to semi-classical transport theory, an electron in a semiconductor drifts under the influence of a macroscopic electric field and experiences occasional random jumps in its momentum due to different scattering mechanisms in the crystal, such as, acoustic and optical phonons, ionized impurities, etc. This motion of an electron can be described by the following stochastic differential equation:

$$\frac{d\vec{x}}{dt} = \vec{v}(\vec{k}) = \frac{1}{\hbar} \nabla_{\vec{k}} \epsilon(\vec{k}) \quad (1)$$

$$\hbar \frac{d\vec{k}}{dt} = -q\vec{E} + \vec{F}_r \quad (2)$$

$$\vec{F}_r = \sum_i \hbar \vec{u}_i \delta(t - t_i) \quad (3)$$

Here, \vec{x} , \vec{v} and \vec{k} are the electron position, drift velocity and wave vector, respectively. \vec{E} is the electric field, $\epsilon(\vec{k})$ is the energy-wave vector relationship and \vec{F}_r is the random impulse force on the electron due to scattering. This random force can be characterized by the scattering rate $\lambda(\vec{k})$ and the transition rate $S(\vec{k}, \vec{k}')$. The probability of scattering in time is given by the following expression:

$$\Pr\{t_i - t_{i-1} > \tau\} = \exp\left\{-\int_{t_{i-1}}^{t_{i-1}+\tau} \lambda(\vec{k}(t')) dt'\right\} \quad (4)$$

Given that a scattering event has occurred at some time t_i , the probability density function for the change in the electron wave-vector is expressed as follows:

$$\rho_{\vec{k}_i}(\vec{u}_i) = \frac{S(\vec{k}_i, \vec{k}_i + \vec{u}_i)}{\lambda(\vec{k}_i)} \quad (5)$$

where $\vec{k}(t_i^-) = \vec{k}_i$ and $\vec{k}(t_i^+) = \vec{k}_i + \vec{u}_i$.

Equations (1)-(5) define a Markov process which is discontinuous in wave-vector space (compound Poisson process). In the theory of stochastic differential equations, such a Markov process is generally characterized by a *transition density function*, which satisfies the Kolmogorov-Feller forward equation, (see Refs. [2] and [3] for further details). By formally integrating this equation over the probability density function of the initial state, one obtains the linear Boltzmann transport equation:

$$\frac{\partial f}{\partial t} + \vec{v} \cdot \nabla_{\vec{x}} f - \frac{q\vec{E}(\vec{x}, t)}{\hbar} \cdot \nabla_{\vec{k}} f = \int f(\vec{x}, \vec{k}', t) S(\vec{k}', \vec{k}) d\vec{k}' - \lambda(\vec{k}) f \quad (6)$$

where f is the electron distribution function.

This establishes a direct connection between semi-classical transport theory and the theory of stochastic differential equations. It also makes it possible to apply the machinery of stochastic differential equations to semiconductor device modeling.

II. The Transport Model:

In this section, we will present a transport model which is based on the Boltzmann transport equation. According to semi-classical transport theory (2), the total force on the electron is due to the electric field \vec{E} and the random scattering force \vec{F}_r . The expected value of this random force can be computed from (3)-(5) and is given by the following expression:

$$\bar{\vec{F}}_r(\vec{k}) = E\{\vec{F}_r\} = \hbar \int \vec{u} S(\vec{k}, \vec{k} + \vec{u}) d\vec{u} \quad (7)$$

This expression is in fact the first moment of the transition rate S . For acoustic and optical phonon scattering, it can be shown that the expected value of the random force is given by the following expression:

$$\bar{\vec{F}}_r(\vec{k}) = -\hbar \lambda(\vec{k}) \vec{k} \quad (8)$$

which can be interpreted as a drag-force opposite in direction to the electron wave-vector. Substituting this expression into (2), one obtains the following set of stochastic differential equations:

$$\frac{d\vec{x}}{dt} = \vec{v}(\vec{k}) \quad (9)$$

$$\hbar \frac{d\vec{k}}{dt} = \left[-q\vec{E}(\vec{x}, t) - \hbar \lambda(\vec{k}) \vec{k} \right] + \vec{F}_r^0 \quad (10)$$

Here, \vec{F}_r^0 denotes the zero mean component of the random (impulse) force.

In order to obtain a transport model suitable for efficient numerical implementation, the integral operator in the Boltzmann transport equation can be approximated by a second order differential operator. This turns out to be equivalent to approximating the zero mean fluctuating force \vec{F}_r^0 by the derivative of a zero mean Wiener process (white noise). The variance of the Wiener process is taken to be equal to the momentum dependent variance of the random force:

$$\sigma_{ij}(\vec{k}) = \int u_i u_j S(\vec{k}, \vec{k} + \vec{u}) d\vec{u} \quad (11)$$

This is in fact the second order moment of the transition rate S . The second order moments of the transition rate can be computed by integrating over equal energy surfaces in \vec{k} -space (spheroids or ellipsoids). For instance, in the case of a spherical band structure, the second moment in the direction orthogonal to \vec{k} is given by the following expression:

$$\sigma_{k_\perp k_\perp}(\vec{k}) = \frac{1}{3} \sum_m \lambda^{(m)}(\vec{k}) r_m^2(\vec{k}; \beta) \quad (12)$$

and, in the direction parallel to \vec{k} , it is given by

$$\sigma_{k_\parallel k_\parallel}(\vec{k}) = \sigma_{k_\perp k_\perp}(\vec{k}) + |\vec{k}|^2 \lambda(\vec{k}) \quad (13)$$

Here, m is the index over different scattering processes, β is the dispersion factor (see Ref. [4]) and r_m is the radius of an equal energy sphere in \vec{k} -space after scattering. Similar relations hold in the case of ellipsoidal bands by the use of the Herring-Vogt transformation (see Ref. [5]). Figs. 1 and 2 show the expected value of the random force $\vec{F}_r(\vec{k})$ and the variances $\sigma_{k_\perp k_\perp}(\vec{k})$ and $\sigma_{k_\parallel k_\parallel}(\vec{k})$, respectively, as a function of the magnitude of the electron wave vector.

Employing the approximation for the zero mean random force results in a second order partial differential equation for the electron distribution function. In spherical coordinates, this equation has the following form:

$$\begin{aligned} \frac{\partial f}{\partial t} + \frac{1}{\hbar} \nabla_{\vec{k}} \varepsilon(\vec{k}) \cdot \nabla_{\vec{x}} f - \nabla_{\vec{k}} \cdot \left(\frac{q \vec{E}(\vec{x}, t) - \vec{F}_r(\vec{k})}{\hbar} \right) f = \\ \frac{1}{2} \left(\frac{\partial^2}{\partial r \partial r} \sigma_{k_\parallel k_\parallel} f + \frac{1}{r^2} \frac{\partial^2}{\partial \theta \partial \theta} \sigma_{k_\perp k_\perp} f + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi \partial \phi} \sigma_{k_\perp k_\perp} f \right) \end{aligned} \quad (14)$$

In the talk, we will present preliminary results on the computation of the electron distribution function employing the above model.

Acknowledgement:

This work was supported by the National Science Foundation Research Initiation Award program.

References:

- [1] C. E. Korman and I. D. Mayergoyz, "Application of Stochastic Differential Equation Theory to Semiconductor Transport," *Proceedings of the International Workshop on Computational Electronics*, pp. 159-162, University of Illinois at Urbana-Champaign, May, 1992.
- [2] I. I. Gihman and A. V. Skorohod, *Stochastic Differential Equations*, (Springer-Verlag, New York, 1972).
- [3] D. Kannan, *An Introduction to Stochastic Processes*, (North Holland, New York, 1979).
- [4] C. Jacobini and L. Reggiani, *Reviews of Modern Phys.*, 55, 645, (1983).
- [5] C. Herring and E. Vogt, *Phys. Rev.*, 101, p. 944, (1956).

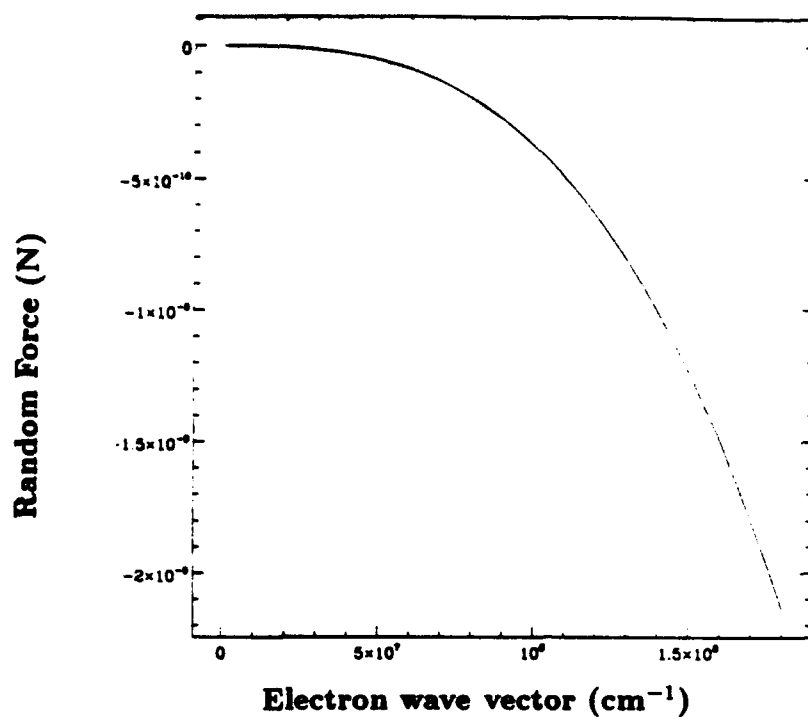


Figure 1. The expected value of the random force vs. the magnitude of the electron wave vector.

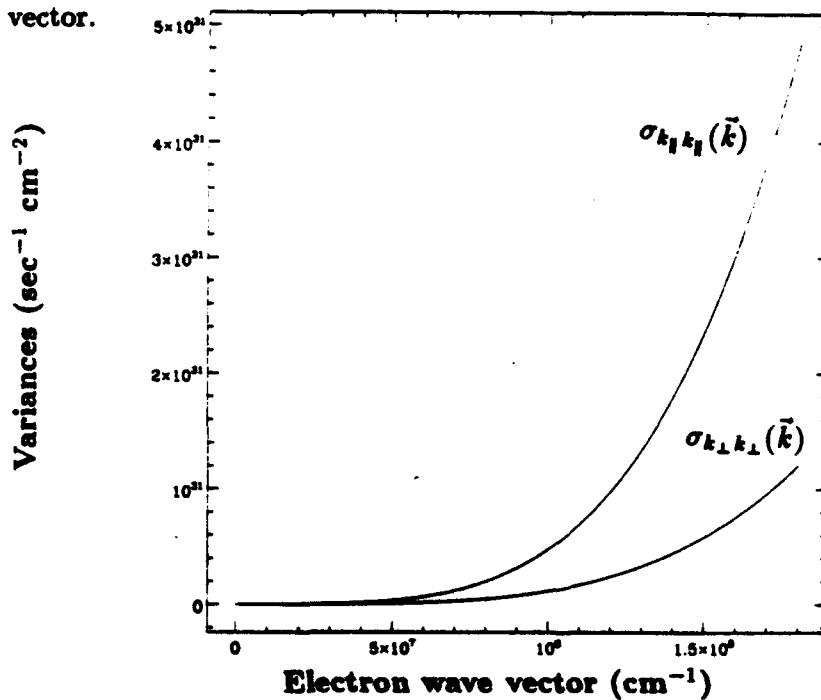


Figure 2. The variances $\sigma_{k_{\perp} k_{\perp}}(\vec{k})$ and $\sigma_{k_{\parallel} k_{\parallel}}(\vec{k})$, respectively, as a function of the magnitude of the electron wave vector.

PMC-3D: A Parallelized 3D Monte Carlo Simulator for Electronic and Electro-optic Devices

S. Pennathur, U.A. Ranawake, V.K. Tripathi, Patrick Lenders and Stephen M. Goodnick

Department of Electrical and Computer Engineering
Oregon State University
Corvallis, OR 97331

Abstract

A parallel Monte Carlo device simulator, PMC-3D has been developed for multiprocessors. Through the use of parallel architectures, full three dimensional modeling of the device domain is possible. Here a discussion of the parallel algorithm is given for coupling the Monte Carlo particle simulation with Poisson's equation for quasi-static problems, and full Maxwell's equations for electro-optic devices.

I. Introduction

PMC-3D is a parallel three dimensional (3-D) Monte Carlo device simulator written for multiprocessors[1][2]. The parallel algorithm is an extension of the standard Monte Carlo device simulation model in 3D in which the particle dynamics generated from the stochastic Monte Carlo method are solved simultaneously with the appropriate set of field equations on a 3D mesh using finite differences. For quasi-static problems such as the behavior of a three terminal MESFET device, Poisson's equation is solved for the potential and corresponding electric field used to drive the particle dynamics. Both the Poisson solver and the Monte Carlo phase are parallelized due to the large computational requirements of solving Poisson's on a 3D grid. The implementation of the Poisson solver is based on an iterative method that uses an odd/even ordering with Chebyshev acceleration. The code was developed both for a distributed memory 1024 node nCUBE multicomputer and a 4-node shared memory Ardent Titan multiprocessor.

II. Parallel Algorithm

In the distributed memory implementation, the spatial domain of the device is subdivided onto separate processors according to a recursive bisection algorithm. Figure 1 shows the decomposition of the MESFET problem for a simple four processor case. The particles and mesh associated with each of the subregions is mapped onto separate processors. The choice of spatial domain size is chosen to roughly balance the number of particles per processor, and adjusts itself throughout the simulation to maintain a balanced processor load. The code is written in C, and contains compiler directives to distribute portions of the code on various processors.

The parallel to sequential speedup was characterized in several ways shown in Fig. 2. The speedups for the fixed size and scaled size problem were measured for both the Poisson phase and the Monte Carlo phase of the code on a 1024 nCUBE hypercube

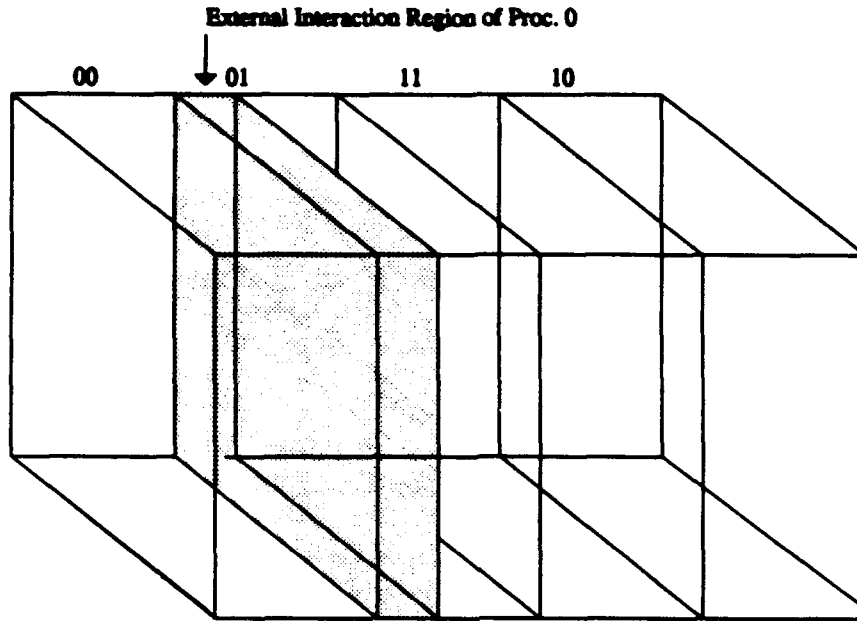


Figure 1: Mapping onto a 4 processor hypercube and external interaction region

with dynamical load balancing turned off. For the fully scaled problem, an efficiency of 69.8% was obtained for 512 processors compared to the ideal (Amdahl's law) speedup. An additional increase of 8% was obtained using dynamical load balancing. In the fully scaled case, several million grid points, and several million particles were simulated.

III. Electromagnetic Modeling

For high frequency operation or for electro-optic devices, Poisson's equation is insufficient to correctly characterize the dynamical behavior of the system, and thus PMC-3D has been extended to include full electromagnetic solutions of Maxwell's equations coupled with the particle transport models. The Lorentz force equation which determines the carrier momentum between collisions, includes the effect of the magnetic and electric fields obtained from the time-dependent solution of Maxwell's equations, and is given by

$$\mathbf{F} = \hbar \dot{\mathbf{k}} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

The standard finite difference/time domain (FD/TD) formulation of the field-equations is used to compute the time-evolving quantities E, H and J with $\rho(x, y, z, t)$ being the source term corresponding to the charges present as described by Connolly *et al.* [3]. The algorithm involves two grids in space (for E and H fields) displaced by half mesh increments in the three spatial directions. Every component of E (or H) is computed at a given time, using the four adjacent values of H (or E) that contribute to the loop integral, that results from the Maxwell's curl equations. In addition, the E and H fields are computed at time instants differing by a half time step, amounting to a leapfrog method of computing the time evolution of the field quantities.

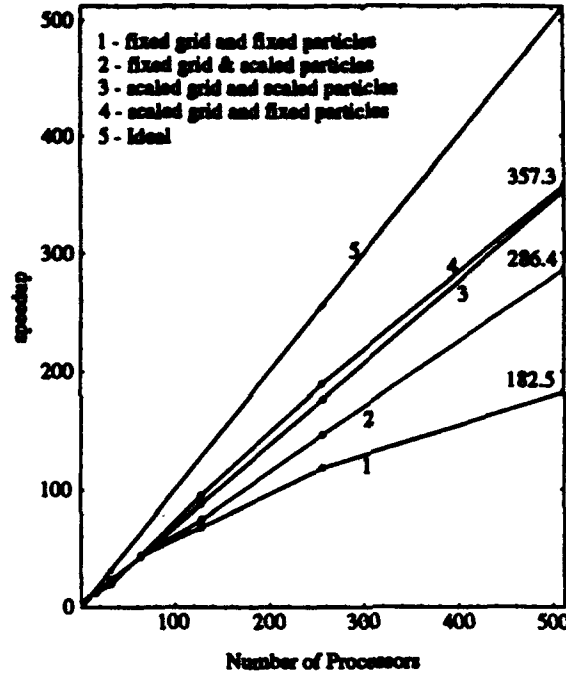


Figure 2: Parallel to sequential speedups of PMC-3D

The current density term J is central to the feedback that exists between the Monte Carlo part and the field computations. It is computed once every timestep at every grid point by summing the weighted velocities of the particles in the ensemble that lie within a unit cell volume about the grid point, and is given as

$$J(i, j, k) = \frac{q}{\Delta x \Delta y \Delta z} \left(\sum_{n=1}^{N(i, j, k)} S_n v_n \right)$$

where S_n and v_n refer respectively to the charge and velocity of the n th particle associated with the grid point [3].

The electric fields are initialized to the values obtained using a one-time solution of Poisson's solution, corresponding to the applied DC bias voltages on the strip line, while the magnetic fields are assumed to be zero. Dirichlet boundary conditions are imposed on the metal-semiconductor boundaries as well as on all the conducting walls. Some preliminary results have also been obtained by imposing Neumann boundary conditions at the semiconductor-air interfaces.

The FD/TD algorithm essentially requires the solution of an identical set of difference equations at every grid point and hence is ideally suited for parallel implementation. In addition to updating the electric and magnetic fields of the grid points belonging to the subgrid mapped to its local memory, every processor also needs to communicate with its neighboring processors via message passing to obtain the electric or magnetic field values that may be required in the computation of fields along the boundaries of subgrids.

The program flow during a typical timestep in each processor starts off with the

computation of electric fields at time t_n (after computing the current density vector J at each grid point at time t_n), followed by evolution of the particles in momentum and real space for half a timestep. After computing the magnetic fields at time $t_n + \frac{\Delta t}{2}$ using the electric field values at t_n , the ensemble is allowed to evolve for another half timestep, thereby completing one timestep of Monte Carlo particle evolution.

One such structure that was modeled is the photoconductive switch [4][5] shown in Fig. 3. As shown there, a femtosecond laser pulse is used to generate electron hole pairs

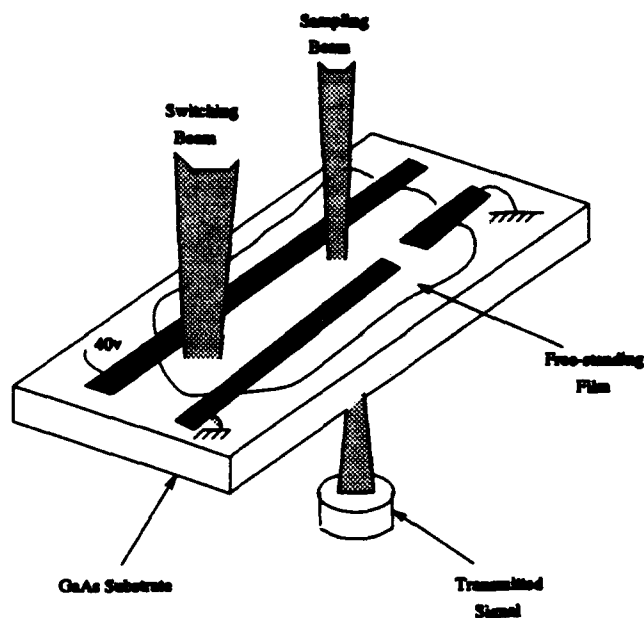


Figure 3: Experimental configuration of photoconductive switching

between two DC biased coplanar striplines on a *GaAs* bulk or superlattice substrate. The transient dynamical response of the electrons and holes as they accelerate in opposite directions induces a subpicosecond time scale transient electrical pulse in the coplanar waveguide which propagates down the stripline and is subsequently detected at a different point. Through this technique, ultra-short time scale electrical pulses may be generated and detected optically. The gap separating the microstrip lines is $10\text{ }\mu\text{m}$, while the strips are $2\text{ }\mu\text{m}$ wide. The *GaAs* semiconducting layer is $1.2\text{ }\mu\text{m}$ thick, and $15\text{ }\mu\text{m}$ long along the z -direction. A 20 femtosecond laser pulse with an energy of 1.55 eV is assumed with a spatial width of $2\text{ }\mu\text{m}$.

Figure 4 shows the simulated particle current for a fairly low (peak) injection density of $1 \times 10^{15} / \text{cm}^3$. The results obtained from using FD/TD solutions of Maxwell's equations as well as the case for which only the Poisson's equation was solved for updating the fields, are shown for comparison. It is seen that there is a reasonable qualitative agreement at low density between the quasi-static solution and the more complete model where velocity overshoot is directly observed. As the density is increased, the perturbation of the DC electric fields due to the higher density of the moving charges results in significant modification of the time dependent magnetic and electric fields. At very high densities, velocity overshoot is barely observed, with additional effects due to reflections of the

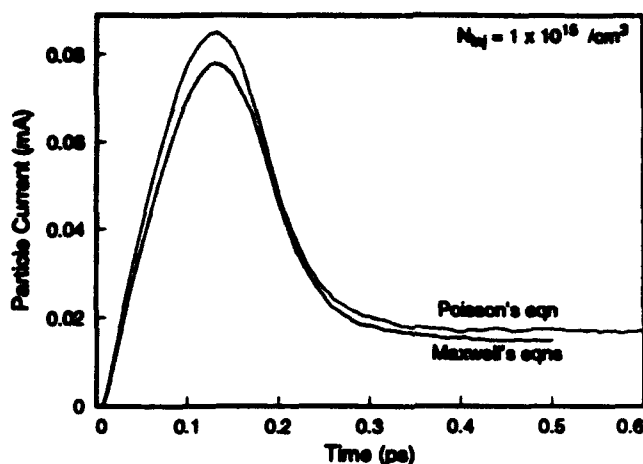


Figure 4: Simulated particle current for $N_i = 1 \times 10^{15} / \text{cm}^3$

electromagnetic fields from the boundary walls coming into play. Comparison of the effects of absorbing versus reflecting boundary conditions are currently being studied.

Acknowledgment

Computer time provided by Sandia National Laboratories and partial support of NSF contract ECS-8821107 are gratefully acknowledged.

References

- [1] U.A. Ranawake, P. Lenders and S.M. Goodnick, "Parallelization of Monte Carlo Algorithms in Semiconductor Device Physics on Hypercube Multiprocessors," *Computational Electronics*, Kluwer Academic Publishers, pp. 137-141, 1991.
- [2] U. A. Ranawake, C. Huster, P. Lenders and S. M. Goodnick, "The 3-D Monte Carlo Simulation of a Semiconductor Device on a Hypercube Multicomputer," Presented at the *Workshop on Computational Electronics*, Urbana, Illinois, May 27-28, 1992; published in the proceedings of the *Workshop on Computational Electronics*, pp. 193-196, 1992.
- [3] K.M. Connolly, R.O. Grondin, R.P. Joshi, and S.M. El-Ghazaly, "Numerical Modeling of Ultrafast Electrical Waveform Generation and Characterization," published in the *Proceedings of Microwave Symposium*, 1990.
- [4] W.H. Knox, J.E. Henry, K.W. Goossen, K.D. Li, B. Tell, D.A.B. Miller, D.S. Chemla, A.C. Gossard, J. English, and S. Schmitt-Rink, "Femtosecond Excitonic Optoelectronics," *IEEE J. Quantum Elec.*, Vol. 25 No. 12 (1989) 2586-2595.
- [5] W.H. Knox, G.E. Doran, J. Cunningham, and S.M. Goodnick, "Femtosecond Spatio-Temporal Field Measurements in GaAs Quantum Wells," in the proceedings of the *Topical Meeting on Picosecond Electronics and Optoelectronics*, Salt Lake City, March 1991 (invited).

Parallel Simulation of Semiconductor Devices on MIMD Machines

Don Reid, Asen Asenov, John R. Barker, Steve P. Beaumont

Nanoelectronics Research Centre,
Department of Electronics and electrical Engineering,
The University of Glasgow,
Glasgow, G12 8QQ, UK

Abstract

In this paper two unified, scalable and transportable parallel approaches to the numerical simulation of semiconductor devices are presented: Concurrent Device Simulation (CDS), and Spatial Device Decomposition (SDD). Both approaches have been developed and tested on the Parsytec Supercluster Model 64 which is a medium size transputer system. A series of examples illustrate the application of the developed parallel simulation tools.

I. Introduction

The rapid progress in device technology, which made possible the fabrication of nanometer scale band engineered structures [1], is now having a considerable impact on device simulation. New physical phenomena such as carrier heating, ballistic transport and quantisation govern the behaviour of nano-scale devices [2] and require more complex and expensive models for simulation. For nano-structure devices, full 3D simulation has becoming a necessity. The need for super computer power often restricts both the complexity of physical models involved in device simulation programs and their widespread application. In the foreseeable future, a significant low-cost improvement in computing performance will only be available through Multiple Instruction Multiple Data (MIMD) systems where necessary speed-up derives from the use of parallel processors sharing a large distributed memory. Hence the parallel implementation of device simulation codes has become essential [3]. Such fast parallel codes could be used for real device optimisation, sensitivity analysis and yield prediction in both research and industry. However, to be of practical use, parallel device simulation programs must be unified, scalable and portable.

We report on two parallel approaches to numerical simulation of semiconductor devices, developed in the Nanoelectronic Research Centre at Glasgow University: Concurrent Device Simulation (CDS), and Spatial Device Decomposition (SDD). Both approaches are implemented on the departmental transputer system - the Parsytec Supercluster Model 64 - which is a medium class MIMD system. It consist of 64 electronically reconfigurable T800 transputers with 4MB of local memory per processor.

II. Concurrent Device Simulation

Any practical simulation run can cover a large matrix of input data parameters from bias points to details of device design such as gate length, vertical layer structure, doping distribution, recess shape etc. A simple but effective form of parallel device simulation is to run several copies of the serial simulation code concurrently, implementing Single Program Multiple Data (SPMD) computational model [4]. To this end a fileserver has been designed to pass input data sets to a pipeline of processors running our serial H2F simulator [5].

The fileserver is split into two parts: the server which is system dependent and the harness which is application dependent. The server runs on a simple pipeline of processors (fig. 1). First it runs a *pre-processing* routine. Then it collects input data packets from a *send-data*

routine, passes them to a free processor on which a *process-data* routine (the simulator H2F in our case) runs, collects data from the *process-data* routine and returns them to the master processor for a *receive-data* routine. This cycle is repeated until all the data packets have been processed and then the server runs a *post-processing* routine.

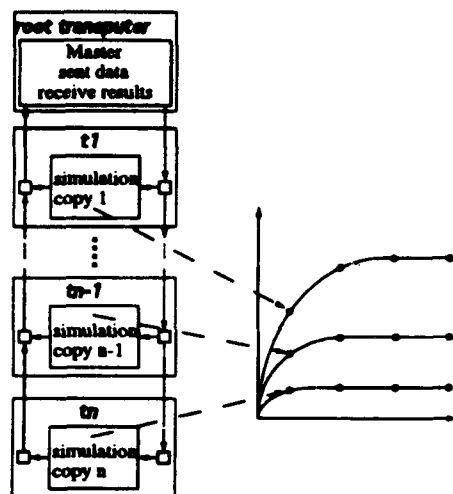


Fig. 1 Pipeline fileserver calculating in parallel an I-V characteristic

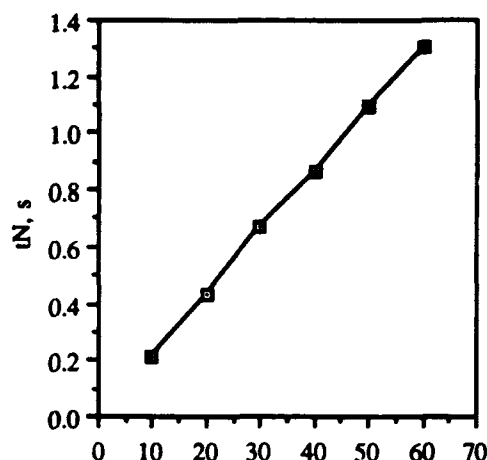


Fig. 2 Time t_N required to pass set of input data to the last processor of an N -processors pipeline

The server is implemented using a simple set of programs and can easily be adapted for a number of platforms e.g. a fast serial processor; Inmos transputer boards running 3L languages; a Parsytec Multiclustor 32/Superclustor 64 running Helios/Parix or a Meiko computing surface of transputers /C40s running CDL. The server links to a harness of five routines - *pre-process*, *process*, *post-process*, *send-data* and *receive-data* which are independent of the system and completely portable. For long computational jobs the fileserver has proved superior to conventional farm processing techniques supported by many of the existing parallel languages because it can be reliably scaled and is not restricted by the platform dependant variations in buffering systems.

The communication time overhead associated with the fileserver is illustrated in Fig. 2 where the measured access time to the last processor is plotted as a function of the fileserver length. Since a single bias point calculation with H2F takes approximately one hour or more, the communication overhead is negligible. However the average access time can be significantly reduced on many of the configurable transputer systems if a ternary tree is used instead of a pipeline. The effectiveness of the CDS when single bias data are calculated on each processor is also restricted by two other factors. First, because the simulation time for different bias conditions can vary significantly, the total execution time is determined by the processor with the worst combination of bias points. Secondly, the techniques of extrapolating the initial guess from previous solutions is inapplicable. The second disadvantage may be reduced by calculating sets of bias points on a each processor.

Investigations of the influence of the structure parameters on the device's performance has proved to be extremely amenable to CDS. Fig. 3 shows a set of I-V characteristics for a 100 nm gate-length δ -doped pseudomorphic HEMT, calculated in parallel with a CDS on 60

processors, with variable δ -doping and gate length. The estimated speed-up for this particular simulation is around 30 which reflects approximately 50% efficiency.

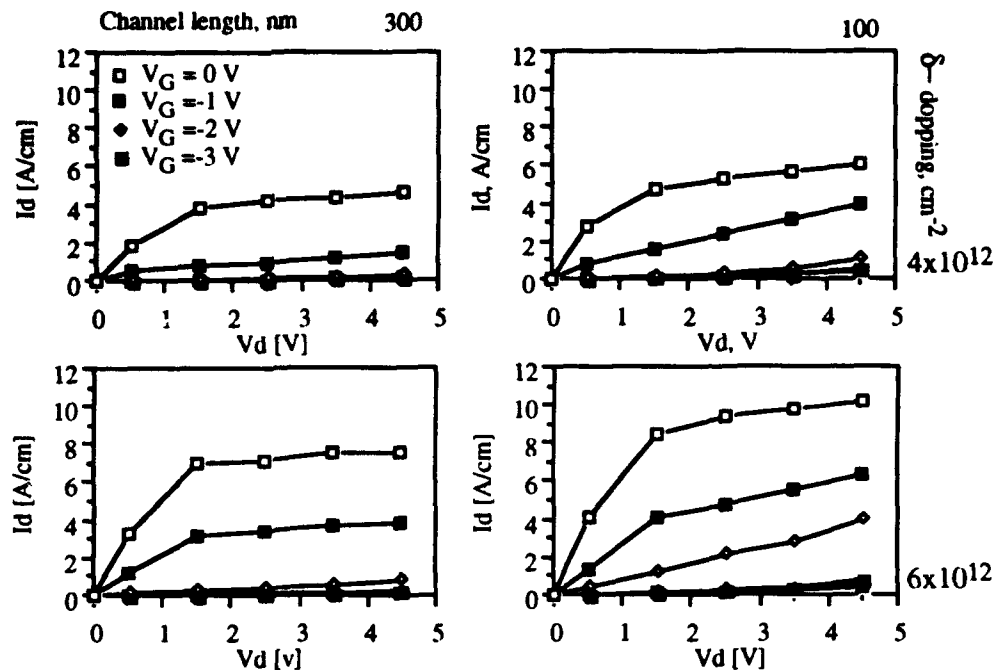


Fig. 3 Set of I-V characteristics for 4 different PHEMTs calculated simultaneously using the CDS approach

III. Spatial Device Decomposition.

The SDD approach is a powerful method of accelerating a single bias point device simulation and of overcoming the inherent memory limitation of 3D simulations. The basic idea of implementing this approach on MIMD is illustrated in Fig. 4 where the spatial device decomposition of a FET and its corresponding grid partitioning is sketched.

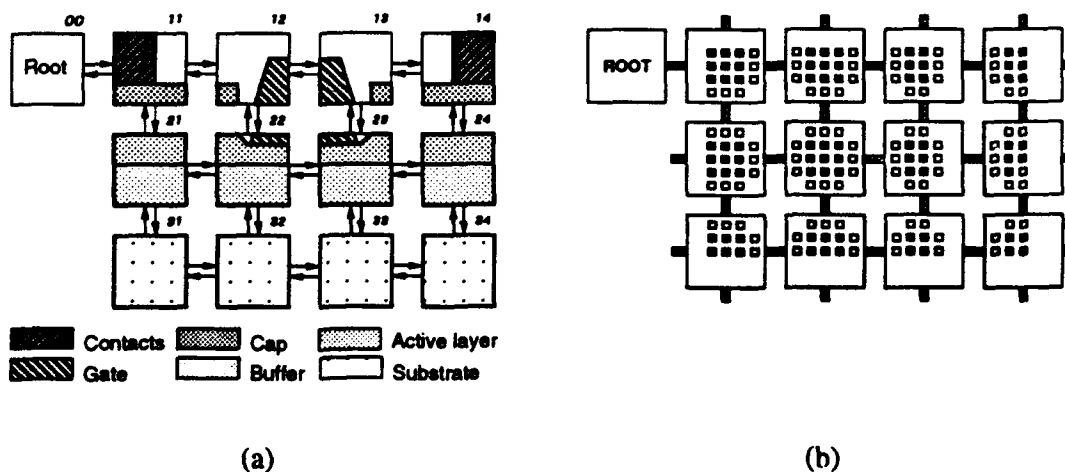


Fig. 4 Spatial device decomposition. (a) physical representation (b) grid partitioning

The solution domain of the device being simulated is spread over a large number of processors (Fig. 4. (a)). If the solution domain is topological rectangular (as in H2F [5]) the best configuration is an array of $N \times M$ processors. In this case each processor is allocated to the grid points corresponding to one device's subdomain plus the grid points from the boundary of the neighbouring processors subdomains (Fig. 4 (b)).

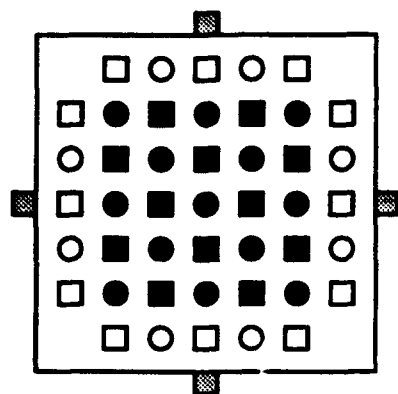


Fig. 5 Implementation of a black/red SOR solver. Black (■) and red (●) nodes are updated simultaneously

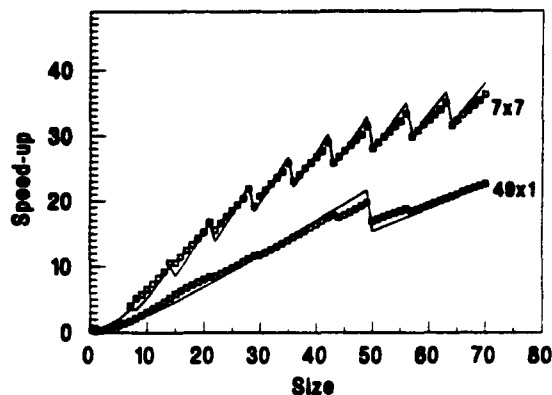


Fig. 6 Speed-up of the parallel SOR solver for an array and pipeline of 49 processors: — theory, □ - measurements

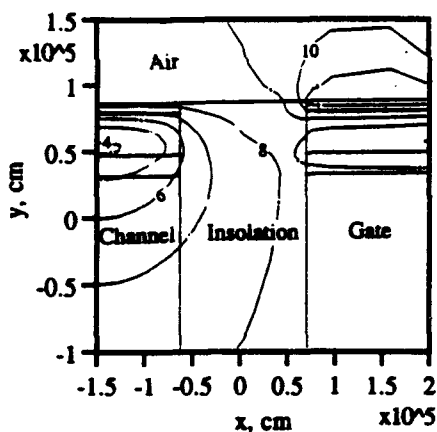


Fig. 7 Potential distribution normal to the channel of an IPGT with gate isolation created by ion implantation

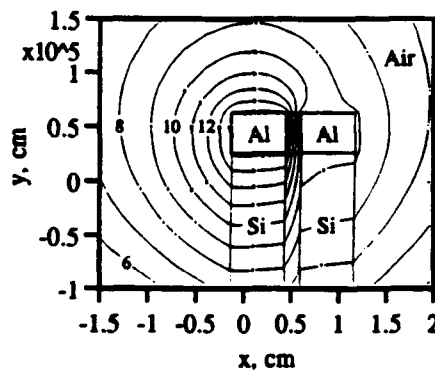


Fig. 8 Potential distribution around two aluminium wires on etched p-Si pedestals for design of Coulomb blockade devices [8]

To ensure portability, our parallel simulation system is divided into two parts: a system dependent communication harness and a system independent simulation harness. The developed communication harness for the Parsytec, GARH (General ARray Harness) provides all global and local communications needed to implement the device simulation code on an array of transputers. The grid generation is performed on the 'root' transputer, then the device is partitioned and distributed over the complete array. All processors perform the data

initialisation and discretization in parallel. The key point is the parallel solution of the algebraic system, arising from the discretization of Poisson's and the current continuity equations. For the discretized Poisson's equation, a black/red SOR solver was implemented (Fig. 5). Applying a recently developed performance theory for the speed-up of the transputer array based iterative solvers [6], the experimentally observed and predicted speed-up is given in Fig. 6. This figure clearly indicates that the array implementation minimises the ratio between the boundary and bulk subdomain grid points and hence the local communications overhead is superior to the pipeline implementation. The development of a processor array current continuity solver is in progress.

Two examples of parallel simulations of nanostructure devices, based only on the parallel solution of Poisson equation are given in Fig. 7 and 8. The first figure represents the potential distribution normal to the channel of an InPlane Gate Transistor (IPGT) [7] with a gate isolation created by ion damaging. The second figure represents the calculation of the capacitive coupling between two aluminium wires on the top of an etched p-Si substrate to aid the design of Coulomb blockade devices [8].

IV. Conclusions

Two unified, scalable and portable approaches have been developed for the intensive simulation of semiconductor devices on medium-sized MIMD systems. The Concurrent Device Simulation approach is simple to implement but leads to a reasonable efficiency only when numerous sets of input data are investigated in the process of device design and optimisation. The Spatial Device Decomposition approach is more complicated but can accelerate a single point simulation. However, the successful implementation of the second approach depends on the development of adequate, effective, scalable and portable linear equation solvers capable of dealing with the difficulties in solving, for example, the discretized current continuity equation.

References

- [1] J. Han, D.K. Ferry and P. Newman, "Ultra-Submicron Gate AlGaAs/GaAs HEMTs", IEEE Electron Dev. Lett., Vol. 11, No. 5, pp.209-211, 1990.
- [2] J.R. Barker and D.K. Ferry, "On the Physics and modelling of small semiconductor devices - I", Solid-State Electron., Vol. 23, No. 6, pp.519-530, 1980.
- [3] J.P. Darling, and I.D. Mayergoyz, "Parallel Algorithms for the Solution of Nonlinear Poisson Equation of Semiconductor Device Theory and Its Implementation on the MPP", J. of Parallel and Distributed Computing, Vol. 8, No. 1, pp.161-168, 1990.
- [4] F. Ferema, D.A. George, V.A. Norton and G.F. Pfister, "A Single-Program-Multiple Data Computational Model for EPEX/FORTRAN", Parallel Computing, Vol. 7, No. 1, pp.11-24, 1988.
- [5] A. Asenov, D. Reid, J.R. Barker, N. Cameron and S.P. Beaumont, "Finite Element Simulation of Recess Gate MESFETs and HEMTs. The Simulator H2F, Accepted for SISDEP, Wien, Austria, 1993.
- [6] A. Asenov, D. Reid, J. R. Barker, "Speed-up of Scalable Iterative Solvers Implemented on an Array of Transputers" Parallel Computing (in press)
- [7] E. Stefanov, A. Asenov, F. Koch, U. Meiners, B.E. Maile, C. Wolk and H. Brugger, "The In-Plane-Gate Transistor: Device Simulation and Design of the IPGT", Extended Abstracts of the 1992 Int. Conf. on Solid State Devices and Materials, Tsukuba, pp.759-761, 1992.
- [8] J.R. Barker, S. Roy and S. Babiker, "Trajectory Representations, Fluctuations, and Stability of Granular Electronic Devices, in Science and Technology of Mesoscopic Structures, Eds. S. Namba, C. Hamaguchi and T. Ando, pp.213-231, 1992.

Numerical Simulation of a GaAs MESFET Device using Parallel Processing Techniques with Adaptive Meshing and Dynamic Load Balancing on a Transputer Network

C.S. Tsang-Ping, D.M. Barry, C.M. Snowden

Department of Electronic & Electrical Engineering, Leeds University, Leeds, LS2 9JT, UK

Summary

A parallel implementation of the numerical simulation of a GaAs MESFET device using the finite difference discretisation scheme and solved by a point iterative method is presented. Parallel techniques targeted at Multiple-Instructions Multiple-Data (MIMD) message-passing distributed memory architectures, in particular, *Transputers*, are described. Efficient parallelism is achieved by the geometric decomposition of the problem domain. Issues on the convergence and efficiency of the solution with novel strategies such as iteration ordering techniques and communication protocols in a parallel environment are discussed. The parallel implementation of the adaptive grid refinement also requires load balancing techniques for optimum efficiency. Performance results of the parallel simulation using these techniques are presented.

1. Introduction

Physical models are now widely used for simulating complex semiconductor devices. The increasing complexity of the device models requires high performance computers particularly for *interactive* device characterisation. With the advent of relatively cheap parallelism in the form of *Transputers* [1], a building block for a multiprocessor MIMD Distributed Memory Parallel System, the computational requirements for fast device characterisation can theoretically be met. Parallel processing offers attractive advantages such as scalable performance and superior cost/performance ratio. While parallel hardware is readily available, there still is a need for robust and efficient parallel software for semiconductor device simulation.

2. The Device Simulation Problem

This work covers the numerical simulation of a typical n-channel Metal Semiconductor Field-Effect Transistor (MESFET) using a physical modelling approach. A simplified *Drift / Diffusion Transport* model with the *Scharfetter-Gummel* formulation for current density has been used. The numerical solution was achieved using the finite difference discretisation scheme. The basic semiconductor equations consisting of the closely coupled poisson and continuity equations are solved by the Gauss-Seidel point iteration method with successive relaxation for time-dependent solution [2]. An adaptive refinement strategy is implemented for numerical accuracy and optimising computer resources.

3. The Parallel Solution - System and Algorithms

The parallel system used consists of an array of 16 TRANputer Modules (TRAMs) which forms a General Purpose MIMD [3] (GP-MIMD) message-passing distributed memory parallel system. A high level portable language, *Parallel ANSI C*, was used to code the algorithms. The device simulation problem has been parallelised using a one dimensional geometric decomposition [4]. Each processor is assigned a slice of the domain which is stored locally on each TRAM module's memory (e.g. Figure 1 shows 4 transputer modules and 4 subdomains). This is easily scalable to n transputers where n is less than the no of grid points horizontally.

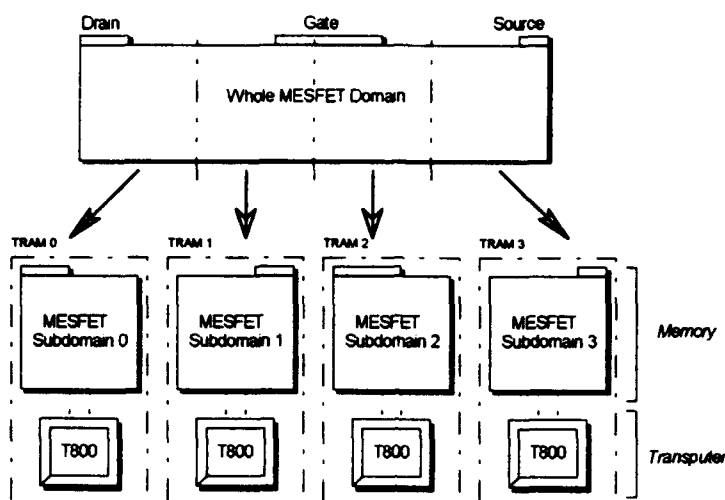


Figure 1. Parallel Decomposition of the Device Simulation Problem

The parallel system configuration of both hardware and software and associated communication channels is shown in figure 2. The transputers are connected in a ring network topology. A graphics process enables the real-time visualisation of the solution.

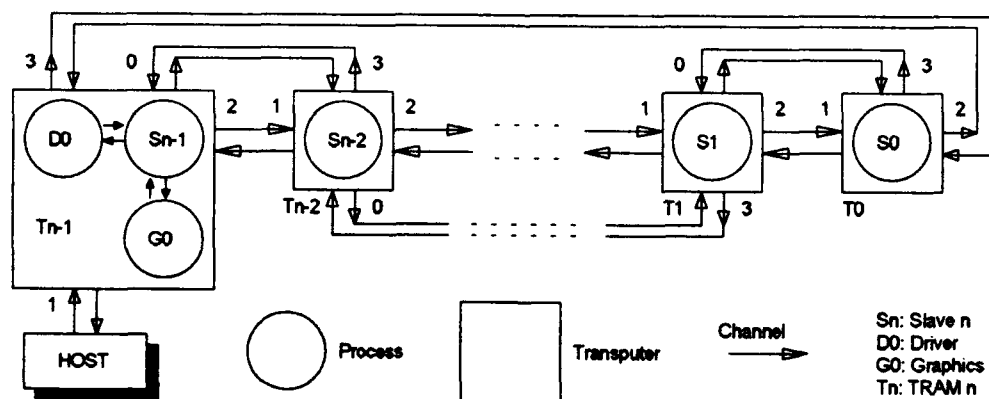


Figure 2. Parallel Hardware and Software Configuration

We now introduce the major overheads in the parallel simulator as opposed to a sequential implementation. The *slave* processes are controlled via channels 1 & 2 (control protocol) by a *driver* process. The latter instructs the slaves to perform specific task and monitors the state of

the simulation such as the global convergence of the iterative solvers and refinement decisions. The five-point formula used for solving the poisson and continuity equations requires four neighbouring data. However, at subdomain boundaries, a column of node data is missing and has to be transmitted via channels 0 & 3 (data exchange protocol).

The decomposition of the *MESFET* domain introduces convergence and stability problems of the numerical solution. For the poisson and continuity iterative solver, the update ordering is modified when domain is partitioned. A novel red/black checker-board updating technique named as the *RB SOR ID* [5] partitioning method that ensures the stability and provides optimal convergence of the numerical solution was developed.

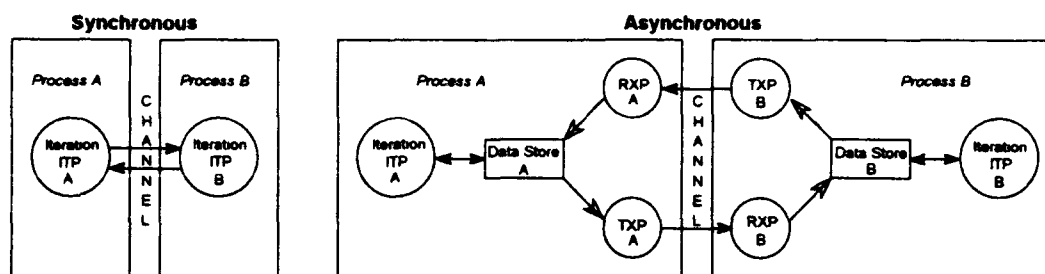


Figure 3. Data Exchange Communication Protocols

Data exchange communication protocol which ensures data consistency across boundaries can broadly be categorised as synchronous and asynchronous [6]. Figure 3 depicts the two types of protocols implemented. Synchronous communication ensures that iteration processes are synchronised and boundary data are updated in a orderly fashion. Asynchronous communications, on the other hand, enable the communication processes (TXP & RXP) to be completely independent of the computation process (ITP). This protocol provides less synchronisation overheads but data consistency at each iteration cannot be predicted.

4. Parallel Adaptive Meshing and Load Balancing

A refinement algorithm based on the potential difference criteria was implemented on the parallel simulator. The dynamic refinement of the domain results in an imbalance of work loads on each processor. Load balancing techniques based on a quasi-dynamic [7] strategy were designed to achieve optimum performance. Figure 4 shows the allocation of the sub-domains when a network of 4 transputers has been load-balanced.

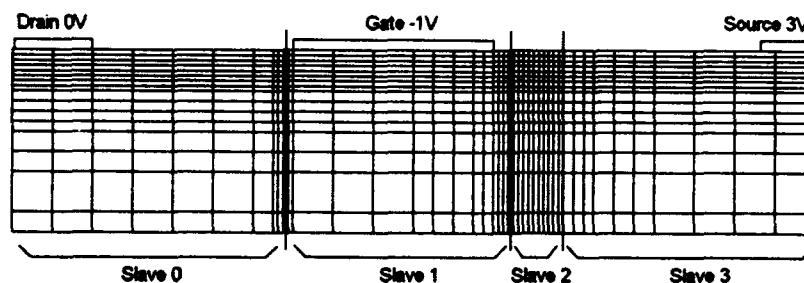


Figure 4. Adaptive meshing and Load Balancing

5. Results and Performance

These results and performance data were obtained from the parallel system by simulating a *GaAs MESFET* with $0.15\mu\text{m}$ active channel ($1.5 \times 10^{23} \text{ m}^{-3}$) and $0.55\mu\text{m}$ gate length using dynamic adaptive grid meshing. The solution for the electron concentration profile is shown in figure 5 for a bias condition shown in figure 4.

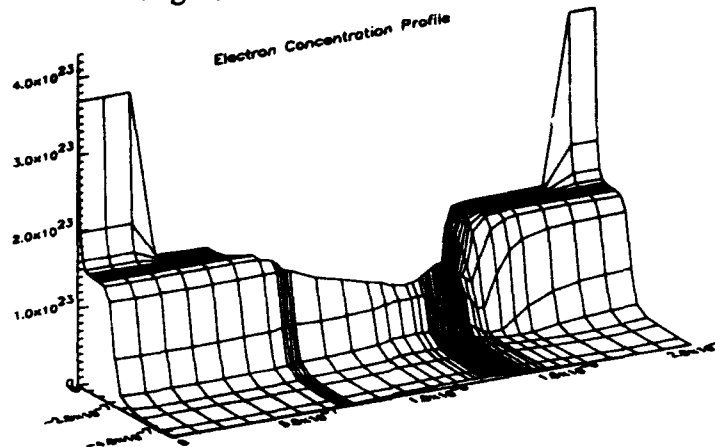


Figure 5. *Electron Concentration Profile for GaAs MESFET*

An overview of the overall performance of the parallel system is shown in figure 6. The drop in speed up is attributed to the communication overheads in the parallel system. The graph shows the parallel performance of the solution for a fixed grid of 128×32 . A speed increase of up to a factor of 12.5 for synchronous protocol on a 16 transputer network has been achieved. For the asynchronous protocol depicted in figure 3, lower performance is observed due to excessive communication overheads.

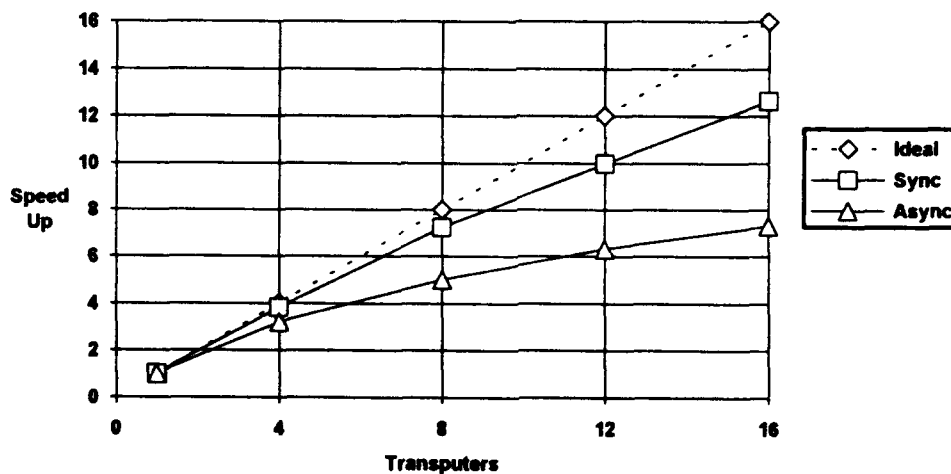


Figure 6. *Overall Parallel Performance Results for fixed grid*

Figure 7 shows the parallel performance for the parallel adaptive meshing algorithm of the *GaAs MESFET* device (from figure 4 & 5) starting with a grid of 41×10 to 62×14 over a time of 50fs. Due to the small grid size, the communication overheads are significant leading to

decreasing performance as more processors are used. The importance of dynamic load balancing is clearly shown from these results (using synchronous and RB SOR 1D techniques).

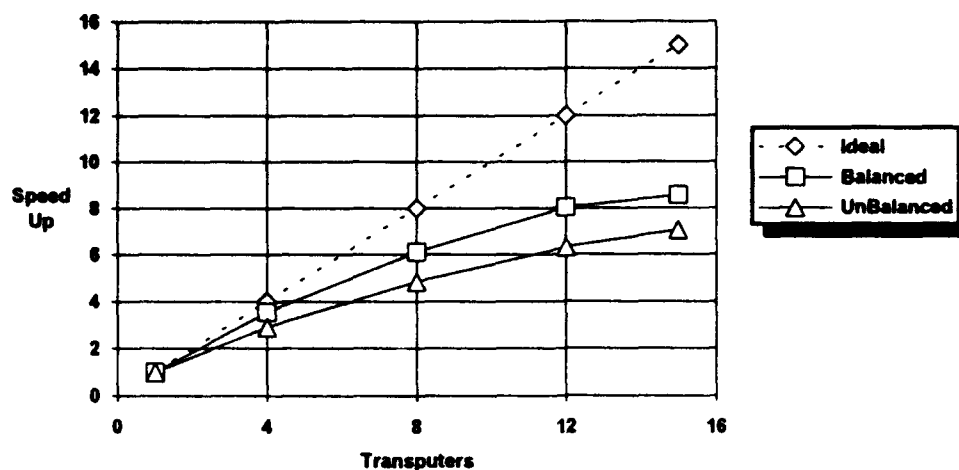


Figure 7. Overall Parallel Performance Results for Adaptive Grid

6. Conclusion

A parallel numerical solver for an iterative GaAs MESFET device solver has been presented. Geometric domain decomposition proves to be the natural way of parallelising the simulation enabling a logical map on transputer-based systems. Combined with Parallel C, this provides a good platform for developing parallel algorithms. However the overall code complexity and size increases in the parallel implementation. The convergence and stability behaviour of the solution is also found to be affected in a parallel environment. This led to the development of suitable iteration update ordering techniques which ensure fast convergence and stability of the solution. Communication overheads remain the major factor limiting the efficiency of the parallel simulator. The parallel grid refinement provided very efficient computation, with the optimum grid structure automatically generated but requires load balancing techniques to improve parallel efficiency. Overall, the performance of the parallel solver is scalable for increasing problem size and numerical complexity. It is envisaged that the parallel techniques developed will have a wide application on exceptionally computer intensive semiconductor device simulation as well as on fast/real-time device characterisation.

References

- [1] A. Trew, G. Wilson, *Past, Present, Parallel: A survey of available parallel computing systems*, Springer-Verlag, 1991.
- [2] T.M. Barton, *Characterisation of the physical behaviour of GaAs MESFETs*, PhD Thesis, Leeds University, 1988.
- [3] M.R. Jane, R.J.Fawcett, T.P. Mawby, "Transputer Applications - Progress and Prospects", *Proceedings of the closing symposium of the SERC/DTI initiative in the engineering apps. of transputers*, IOS Press, 1992.
- [4] C.G. Fox, M. Johnson, G. Lyzenga, S. Otto, J. Salmon, D.Walker, *Solving problems on concurrent processors: General techniques and regular problems*, Prentice-Hall, 1988.
- [5] C.S. Tsang-Ping, D.M. Barry, C.M. Snowden, "Application of Parallel Processing Techniques to GaAs MESFET Device Simulation", *Proceedings of the 4th GaAs Simulation Group Meeting*, Harrogate, UK
- [6] D.P. Bertsekas, J.N. Tsitsiklis, *Parallel and Distributed Computation*, Prentice Hall Inc., 1989
- [7] R.D. Williams, "Performance of dynamic load balancing algorithms for unstructured mesh calculations", *Concurrency: Practice and Experience*, Vol.3(5), pp. 457-481, Oct 1991.

The Implementation and Speed-up of Coloured SOR Methods for Solving the 3D Poisson Equation on an Array of Transputers

Andrew R. Brown, Don Reid, Asen Asenov and John R. Barker

Nanoelectronics Research Centre,
Department of Electronics and Electrical Engineering,
University of Glasgow,
Glasgow, G12 8QQ, UK

Abstract

In this paper we present four different SOR variants for solving the 3D Poisson equation on an array of transputers. The performance of all variants has been tested and compared in terms of number of iterations and global computation time for different configurations of the transputer system. The parallel performance of the algorithms has also been evaluated and compared to a theoretical speed-up model.

I. Introduction

The realistic semiconductor device simulation (both classical, Monte Carlo or quantum mechanical) in many cases requires a 3D solution of the Poisson equation and leads to enormous problem sizes [1]. The single processor implementation of the corresponding 3D codes is limited by both the processor speed and the huge memory-access bottleneck. In the foreseeable future a significant low cost improvement in computer performance will only be available through Multiple Instruction Multiple Data (MIMD) systems (many of them transputer based), for which the necessary speed-up derives from the use of parallel processors sharing a large distributed memory. The point and block Successive Over Relaxation (SOR) methods are promising candidates for 3D parallel implementation on such computers. Although the recursive character of the original SOR method seems to be a serious impediment [2], for a large class of linear systems arising from finite difference, and in particular cases [3] from finite element approximations of the Poisson equation, the multicolour ordering of the grid points leads to easily parallelizable versions of the SOR method [4,5].

Here we present a systematic approach to the parallel implementation of scalable point and block black/red 3D SOR Poisson solvers on a 2D arrays of transputers for the purposes of semiconductor device simulation. Utilising the power and the flexibility of the Parsytec Supercluster Model 64 a wide range of experiments have been made both to compare the performance of the different SOR variants and to choose the optimum 2D transputer configuration for mapping 3D problems. The recently developed detailed performance theory [6] has been applied to underpin the experimental solver design.

II. Model Problem and Partition

The Poisson equation used in the semiconductor device simulations

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = -\frac{\rho}{\epsilon} \quad (1)$$

may be nonlinear or linear, depending on whether the charge density ρ is a function of the electrostatic potential ψ or not. In the drift-diffusion and the hydrodynamic simulation approaches it is mainly used in a nonlinear form. In Monte Carlo and quantum mechanical simulation however it is usually enough to solve the linear Poisson equation.

In this study for simplicity and clarity we consider the linear 3D Poisson equation in a unit cube Ω with Dirichlet boundary conditions. A uniform grid is used to discretize the region Ω into Ω_h . The approximation of the second-order derivatives by the second-order accurate central differences on Ω_h leads to a set of $n \times n \times n$ algebraic equations

$$\psi_{i-1,j,k} + \psi_{i+1,j,k} + \psi_{i,j-1,k} + \psi_{i,j+1,k} + \psi_{i,j,k-1} + \psi_{i,j,k+1} - 6\psi_{i,j,k} = b \frac{\rho_{i,j,k}}{\epsilon} \quad (2)$$

for $i, j, k = 2, 3, \dots, n-1$, where n is the number of points on a side of Ω_h , and b is a coefficient depending on n . In most of the following experiments ψ was set to one, ρ was set to zero and the initial conditions were $\bar{\psi} = 0$.

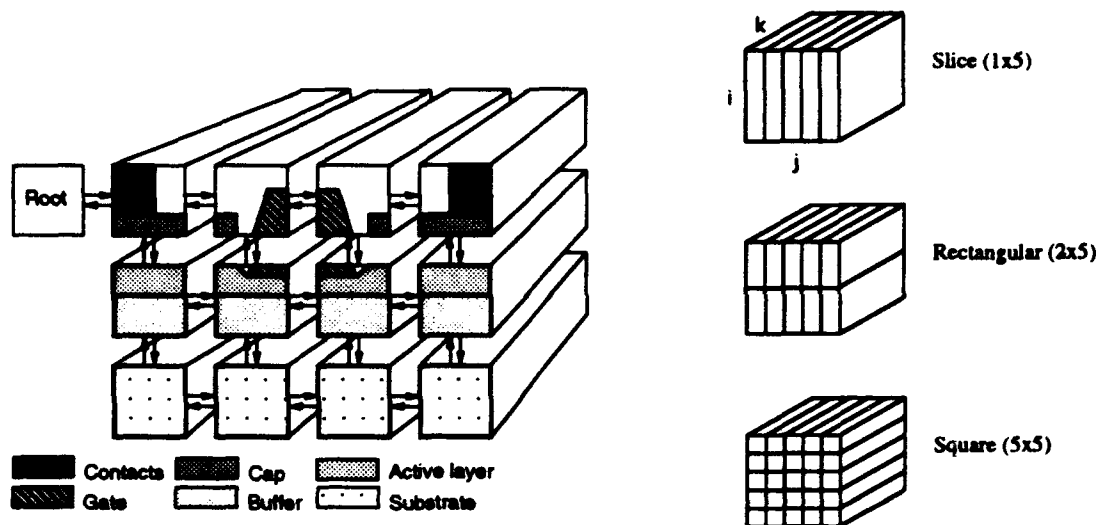


Fig. 1 Mapping of a 3D semiconductor device simulation domain on an 2D array of transputers

Fig. 2 Partition of a 2D discretization grid on a 2D array of transputers

Although it is intuitively clear that the best environment for solving topologically rectangular 3D problems is a 3D array of processors, we are restrict to a 2D array of $N \times M$ transputers. This reflects the connectivity of transputers, which have only four links, and offers a simple way of mapping a topologically rectangular 3D grid, where the overlap of subdomain boundaries is essential. The transputers have a natural ordering $p = 1, 2, \dots, N$ and $q = 1, 2, \dots, M$ Fig. 1. One additional 'root' transputer performs all management and synchronisation. The grid is partitioned into $N \times M$ subdomains along two of the spatial dimensions (i, j) and each of the subdomains involves all corresponding points in the third dimension k . Examples of slice ($1 \times M$), rectangular ($N \times M$) and square ($N \times N$) partitions are given in Fig. 2. A universal communication harness GARH [6] supports all necessary global and local communications.

III. Back/Red SOR Variants

The implementation of the Point SOR method with Natural ordering in i, j directions on each processor may in many cases cause divergence for problems which behave well in a serial implementation [7]. This may be avoided if a black/red SOR variant is considered. In this approach each partition sub domain $\Omega_h^{p,q}$ is decomposed into two further subdomains - black

$\Omega_h^{Bp,q}$ and red $\Omega_h^{Rp,q}$. The nodes in the black and red subdomains are updated simultaneously and the overlapping boundaries are exchanged before each updating.

When a 2D array of processors is used, it is enough to apply the black/red ordering only in the i,j plane which leaves some degree of freedom in the k direction. To explore this freedom we consider four different parallel SOR schemes namely:

PSORBR_{ij}N_k - Point SOR with Black/Red ordering in i,j direction and Natural ordering in k direction. (Fig. 3 a)

PSORBR_{ijk} - Point SOR with Black/Red ordering in all three directions (Fig. 3 b).

PSORBR_{ij}A_k - Point SOR with Black/Red ordering in i,j direction and Alternating directions in k direction.

BSORBR_{ij}T_k - Block SOR with Black/Red ordering in i,j direction and tridiagonal equations solution in k direction.

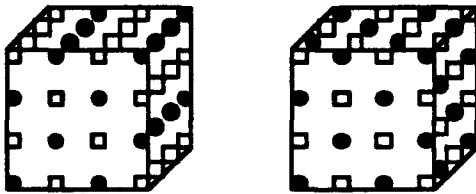


Fig. 3 Black/red ordering in the 3D case.
(a) full black/red ordering (b) black/red ordering only in i,j plane

Table 1

Mesh size n	Natural ordering theory	PSORBR	PSORBRN	PSORBRA	PSORBRT
10	16	15	15	15	11
20	32	32	32	30	22
30	47	47	47	43	33
40	62	61	61	56	41
50	76	74	74	70	51

Table 2

Mesh size n	PSORBR with ω_e	PSORBRN with ω_h	PSORBRN with ω_e	PSORBRA with ω_e	BSORBRT with ω_h	BSORBRT with ω_e
10	0.087	0.093	0.087	0.087	0.106	0.077
20	0.635	0.635	0.635	0.595	0.677	0.552
30	2.184	2.184	2.184	1.998	2.346	1.985
40	5.450	5.539	5.450	5.003	5.888	4.828
50	11.278	11.582	11.278	10.668	12.163	10.169

The main disadvantage of the black/red ordering with all variants described above is that the natural ordering SOR theory [7] does not hold, which makes it difficult to estimate the optimum relaxation coefficient ω a priori. In order to compare more precisely our four different SOR variants, experimental optimum values for ω_e have been found by using a one dimensional minimum search based on the Golden Section method. In Table 1 the numbers of iterations providing an accuracy $\delta=0.001$ are given for both ω_e and the theoretically predicted

relaxation coefficients ω_b [7]. As can be expected the block variant BSORBRT leads to a significant reduction in the number of iterations but at the expenses of an increasing number of calculations per grid point. This point becomes clear from Table. 2 where the execution times in seconds are compared when all 64 transputers of a Parsytec Super Cluster were used, configured in an 8x8 array.

IV Speed-up analysis

To evaluate the parallel potential of the considered SOR variant and the optimal processor configuration we use the relative speed-up defined as the ratio between execution times on an array of processor and the execution time on a single processor. It is obvious that according to this definition the maximal available speed-up is equal to the number of processors on which the algorithm runs. The speed-up model developed in [6] for scalable 2D linear solvers implemented on an array of transputers was extended to the 3D case and takes into account the details of the solution domain partition, all global and local communication overheads and the computation time in the linear and nonlinear case. All parameters of the performance theory have been extracted from independent measurements.

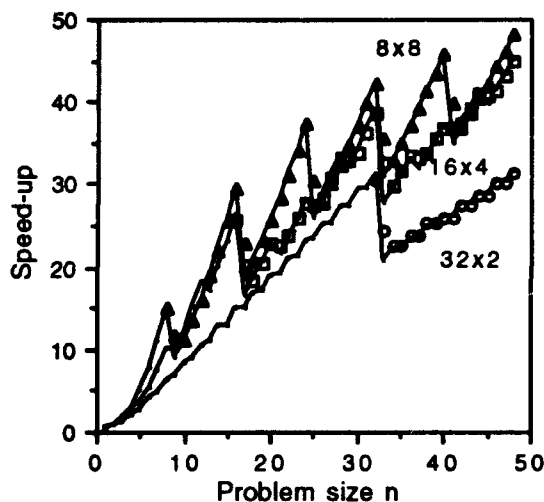


Fig. 4 Speed-up as a function of the problem size for different configurations of 64 transputers (PSORBRN variant)

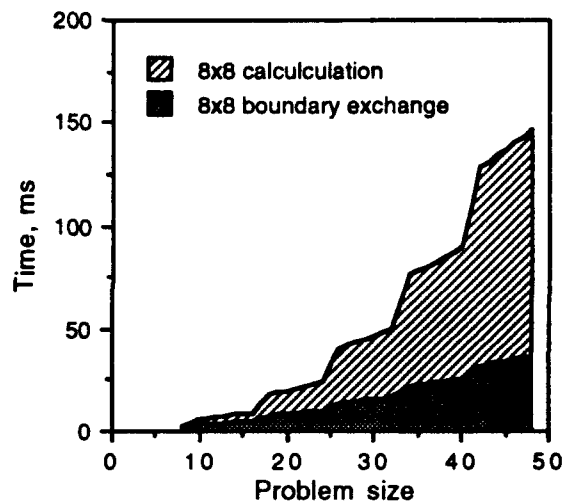


Fig. 5 Balance between calculation time and local communication time in the case of 8x8 transputers (PSORBRN)

The measured and calculated speed-up for three different configurations of 64 transputers is given in Fig. 4 for the PSORBRN variant as a function of the problem size. The picture shows very good agreement between the measured and predicted performance even in the finest details. The square partition which minimises the local communications shows the best performance for problem sizes n which are divisible by the transputer array size. Fig. 5 gives an idea for the balance between the local communication and calculation time for a single iteration in the 8x8 case. Finally Fig. 8 illustrates how the performance theory may be used to estimate the expected speed up for a particular problem size ($n=50$) mapped on a large number of transputers. It is clear that for large transputer systems the BSORBRT variant is superior to all other variants because of the higher calculations/ communications ratio.

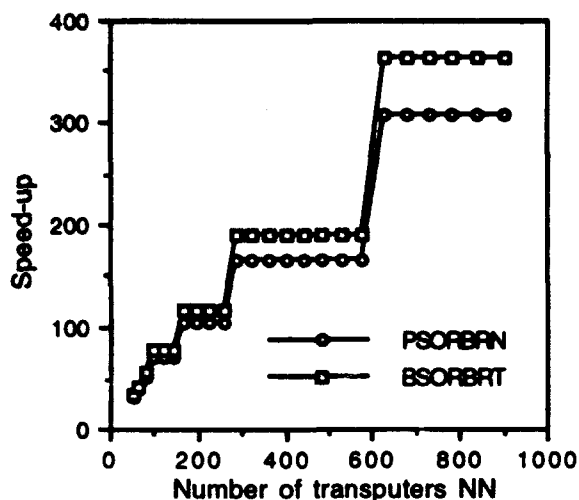


Fig. 6 Speed-up of 50x50x50 problem on a large array of transputers

V. Conclusions

Four variants of a black/red SOR method for solving the 3D Poisson equation on array of transputers have been implemented and tested. The behaviour of PSORBR and PSORBRN variants are very similar in terms of iterations required to achieve a certain accuracy. The third point variant PSORBRA slightly reduces the number of iterations. The block BSORBRT variant reduces the number of iteration by more than 25% but at the expense of a larger calculation time per iteration. Although for a medium size transputer system all four variants are very similar in terms of global computational time, our speed-up analysis shows that on a large array of transputers the advantages of BSORBRT will become more pronounced.

References

- [1] H.C. Chan and T-J. Shien, "A Three-Dimensional Semiconductor Device Simulator for GaAs/AlGaAs Heterojunction Bipolar Transistor Analysis," IEEE Trans. Electron Dev., Vol. 38, No. 11, pp.2427 - 2432, 1991.
- [2] L. Adams and J. Ortega, "A multicolour SOR method for parallel computation," in Proc. 1982 International Conference on Parallel Processing, Bellaire, MI pp.53-56, 1982.
- [3] A. Asenov, D. Reid, J.R. Barker, N. Cameron and S.P. Beaumont, "Application of quadrilateral finite elements for simulation of recess T-gate MESFETs and HEMTs", in this Preceedings, 1993.
- [4] D.J. Evans, "Parallel S.O.R. iterative methods", Parallel Computing, Vol. 1, No. 1, pp.3-18, 1984.
- [5] U. Block, "Block colouring schemes for SOR method on local memory parallel computers", Parallel Computing, Vol. 14, No. 1, pp.61-75, 1990.
- [6] A. Asenov, D. Reid and J.R. Barker, Speed-up of scalable iterative solvers implemented on an array of transputers, Parallel Computing, in press.
- [7] D. Kamowitz, "SOR and MGR[v] experiments on the Crystal multicomputer", Parallel Computing, Vol. 4, No. 2, pp.117-142, 1987.
- [8] R.S. Varga, "Matrix Iterative Analysis", Printice-Hall, Englewood Cliffs, NJ, 1962.

First Principles Calculation of Electron-Phonon Scattering Rates in Si

P. D. Yoder

Beckman Institute for Advanced Science and Technology, and
Coordinated Science Laboratory, Urbana, IL 61801 USA

Abstract

A first principles LDA approach based on the Harris functional ansatz is used to investigate the electron-phonon interaction in silicon, within the rigid ion approximation. Electron-phonon matrix elements for transitions between selected electronic states are calculated, and used to generate band- and wavevector-dependent scattering rates. These scattering rates are generally larger than those calculated with other methods, and show significant anisotropy.

I. Introduction

The concept of applying pseudopotential to the analysis of electron-phonon interactions in solids [1] has served as an invaluable tool in our theoretical understanding of the process. Without pseudopotentials, one currently either relies on a restricted scope of measurable information, or is relegated to phenomenology. Primarily for purposes of simplicity, most investigations of the electron-phonon interaction in semiconductors have made use of empirical pseudopotentials [2-6], mostly in the local approximation [2-5]. As the true crystalline pseudopotential is necessarily nonlocal, some authors have added correction terms [5,6]. While useful in gaining insight, these empirical approaches suffer from their sensitivity to several adjustable parameters as some authors have noted [5, 7]. For example, before any electron-phonon analysis takes place, all of the local form factors must be treated as free parameters and adjusted such that the resulting bandstructure matches experimental values at a suitably defined set of points in the Brillouin zone [8]. Additionally, due to the nature of the electron-phonon interaction potential, one must still choose among various interpolation and extrapolation schemes between and beyond these discrete adjustable form factors in reciprocal space. Of particular concern is the finding that the most physically meaningful choice of extrapolation criteria do not yield the best results [5]. For these and other reasons, several authors have chosen an alternative approach.

It has long been known [9] that electron-phonon matrix elements calculated using pseudopotentials can in principle be just as accurate as those calculated using the true potential, as long as the pseudopotentials are calculated directly from the true potential. This has led to several *ab initio* investigations based on density functional theory [10, 11, 12] in the local density approximation (LDA). Such self-consistent Kohn-Sham (K-S) type calculations quite naturally avoid all of the difficulties of the empirical approach mentioned above. Furthermore, the self-consistent LDA calculations implicitly include the effects of atomic polarization, which is impossible to include using the empirical pseudopotentials.

With all of these benefits, the K-S method has the drawback that modern computational resources limit its application to a small set of high-symmetry transitions due to the requirement of self-consistency.

An alternative approach [7] which uses the *ab initio* pseudopotentials yet avoids the problem of self-consistency involves an approximation to the change in crystalline charge density when the atoms are displaced. This approximation, based on the Harris functional ansatz [13], is to replace the self-consistent K-S potential with one derived from overlapping atomic charge densities. The benefit of this approach is the flexibility to calculate electron-phonon matrix elements for arbitrary transitions, yet it remains a first-principles theory. In the following sections, the method is briefly described, and then used to generate band- and wavevector-dependent scattering rates for silicon.

II. The Model

One may write a crystalline density functional non-interacting Hamiltonian using pseudopotentials as

$$H = -\frac{1}{2}\nabla^2 + \sum_l \Delta V_{ps,nl}^l(\mathbf{r}, \mathbf{r}') + V_{ps,loc}(\mathbf{r}) + V_H[n(\mathbf{r})] + V_{ex}[n(\mathbf{r})] + V_{co}[n(\mathbf{r})], \quad (1)$$

where $V_{ps,loc}(\mathbf{r})$ is the local crystalline ionic pseudopotential, $\Delta V_{ps,nl}^l(\mathbf{r}, \mathbf{r}')$ is the nonlocal l -dependent ionic crystalline pseudopotential, and $V_H[n(\mathbf{r})]$ is the Hartree potential. $V_{ex}[n(\mathbf{r})]$ and $V_{co}[n(\mathbf{r})]$ are functional derivatives of the exchange and correlation energies with respect to pseudocharge density. Single electron wavefunctions are expanded in a plane wave basis as

$$\Psi_{\mathbf{k},n}(\mathbf{r}) = \frac{1}{\sqrt{N\Omega_0}} \sum_{\mathbf{G}} Z_{\mathbf{k},n}(\mathbf{G}) e^{i(\mathbf{G}+\mathbf{k})\cdot\mathbf{r}}, \quad (2)$$

and matrix elements of this pseudopotential Hamiltonian are taken with respect to these basis functions leading to the familiar secular equation

$$\det \left| \left(\frac{1}{2}(\mathbf{k} + \mathbf{G}')^2 - E_{\mathbf{k},n} \right) \delta_{\mathbf{G}',\mathbf{G}} + V_{ps,tot}(\mathbf{k} + \mathbf{G}, \mathbf{k} + \mathbf{G}') \right| = 0 \quad (3)$$

which defines the eigen-energies and wavefunctions.

As is usual, the bare ionic potential (nucleus with core electrons) is expanded in a series under the assumption of small displacements,

$$V_b(\mathbf{r}) = \sum_{j,\alpha} [V_b(\mathbf{r} - \mathbf{R}_j - \tau_\alpha) + \delta \mathbf{R}_{j,\alpha} \cdot \nabla V_b^a(\mathbf{r} - \mathbf{R}_j - \tau_\alpha) + \dots] , \quad (4)$$

ignoring multi-phonon processes [14] and any polarization of the bare ions themselves under displacement. Since the first order term is much smaller than the zero order term, the electron-phonon interaction is treated perturbatively, with the screened first order term giving rise to transitions between eigenstates of the screened zero order term. In the

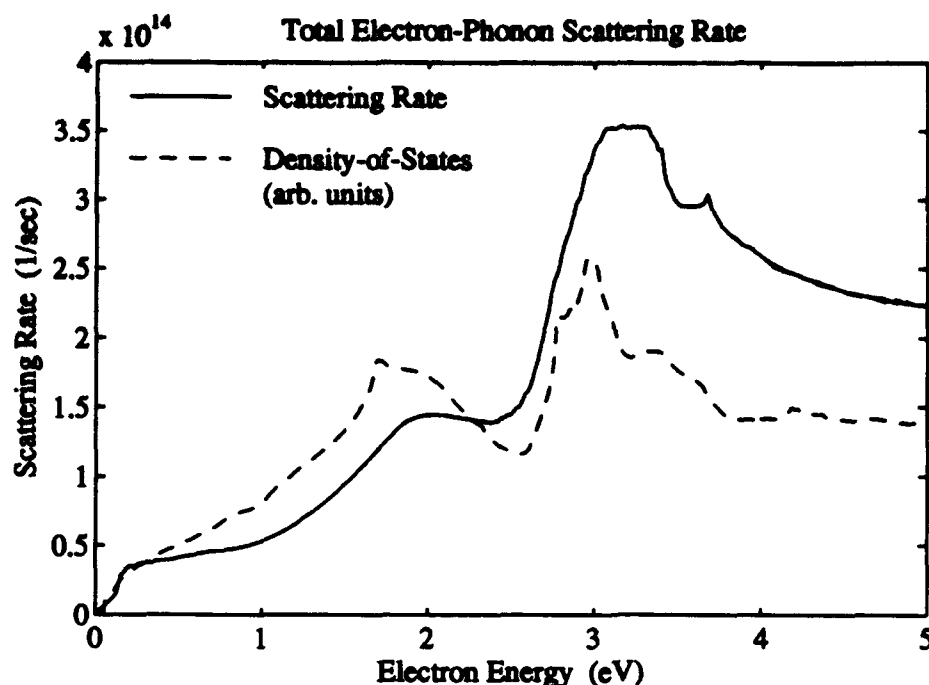


Fig. 1. Total electron-phonon scattering rate as a function of energy. Rates are into states in the first 4 conduction bands out of the first two.

adiabatic approximation, this is expressed as:

$$V_{el-ph}(\mathbf{r}) = \int d^3\mathbf{r}' \epsilon^{-1}(\mathbf{r}, \mathbf{r}'; \omega = 0) \sum_{j,\alpha} \delta \mathbf{R}_{j,\alpha} \cdot \nabla V_b^a(\mathbf{r}' - \mathbf{R}_j - \tau_\alpha) \quad (5)$$

with transform [14]

$$V_{el-ph}(\mathbf{q} + \mathbf{G}) = \sum_{\mathbf{G}'} \epsilon^{-1}(\mathbf{q} + \mathbf{G}, \mathbf{q} + \mathbf{G}'; \omega = 0) V_{el-ph,b}(\mathbf{q} + \mathbf{G}'), \quad (6)$$

defining

$$V_{el-ph,b}(\mathbf{r}) = \sum_{j,\alpha} \delta \mathbf{R}_{j,\alpha} \cdot \nabla V_b^a(\mathbf{r} - \mathbf{R}_j - \tau_\alpha) \quad (7)$$

Under the assumption that the atomic pseudocharge density moves rigidly with the atoms of the crystal lattice under displacement, a linearized interaction potential reduces eq'n (5) to the inner product of the ionic displacement and the gradient of the screened ionic potential. In this case, the calculation of electron-phonon matrix elements proceeds closely along the lines of the standard rigid ion theory, with only minor modification.

III. Discussion

The method of [15] was used in conjunction with the Fermi Golden Rule and a realistic phonon dispersion to generate band- and wavevector-dependent electron-phonon scattering rates. Figure 1 shows the total energy-dependent scattering rate, with the density-

of-states included as a reference. The scattering rate is found to be larger than that of the empirical method [5], especially at high energies. As an indication of the anisotropy of these rates, figure 2 depicts an energy isosurface in the first conduction band, shaded proportionally to total rate at each point in the Brillouin zone.

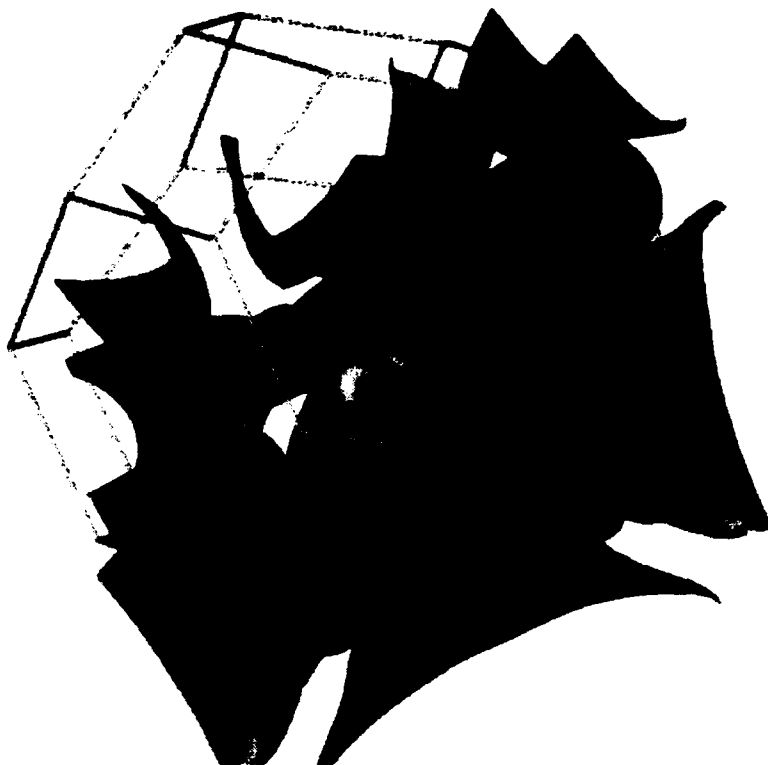


Fig. 2. Surfaces of constant energy ($=1.79$ eV), shaded by electron-phonon scattering rate in the lower half of the Brillouin zone. On these surfaces, the scattering rate varies by a factor of 3.94, indicating significant anisotropy.

Acknowledgments

This research was supported by the Army Research Office. Computer time donated by Cray Research is gratefully acknowledged. The author would like to thank Dr. Karl Hess for his guidance, Dr. P. Vogl and Dr. J. Higman for useful electron-phonon discussions, and Lars Bishop for graphics visualization expertise.

References

- [1] J. M. Ziman, "A Theory of the Electrical Properties of Liquid Metals I: The Mono-valent Metals," *Philos. Mag.* 6, 1013 (1961).

- [2] O. J. Glembocki and F. H. Pollack, "Calculation of the $\Gamma - \Delta$ Electron-Phonon and Hole-Phonon Scattering Matrix Elements in Silicon," *Phys. Rev. Lett.* **48**, 413 (1982).
- [3] S. Zollner, S. Gopalan and M. Cardona, "Microscopic theory of intervalley scattering in GaAs: k dependence of deformation potentials and scattering rates," *J. Appl. Phys.* **68**, 1682 (1990).
- [4] M. L. Cohen and Y. W. Tsang, "Theory of the electronic structure of some IV-VI semiconductors," in *Proceedings of the Conference of Physics of Semimetals and Narrow-Gap Semiconductors*, edited by D. L. Carter and R. T. Bate (Pergamon, New York, 1971).
- [5] M. V. Fischetti and J. M. Higman, "Theory and Calculation of the Deformation Potential Electron-Phonon Scattering Rates in Semiconductors," in *Monte Carlo Device Simulation: Full Band and Beyond*, edited by K. Hess (Kluwer, Dordrecht, 1991).
- [6] D. C. Herbert, "Electron-phonon interaction and inter-valley scattering in semiconductors," *J. Phys. C* **6**, 2788 (1973).
- [7] P. D. Yoder, V. D. Natoli and Richard M. Martin, "Ab initio analysis of the electron-phonon interaction in silicon," *J. Appl. Phys.* **73**, 4378 (1993).
- [8] J. R. Chelikowsky and M. L. Cohen, "Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors," *Phys. Rev. B* **14**, 556 (1976).
- [9] L. J. Sham, "Electron-Phonon Interaction by the Method of Pseudo-potentials," *Proc. Phys. Soc* **78**, 895 (1961).
- [10] O. H. Nielsen and R. M. Martin, "Stresses in semiconductors: Ab initio calculations on Si, Ge, and GaAs," *Phys. Rev. B* **32**, 3792 (1985).
- [11] Z-Q Gu, M-F Li, J-Q Wang and B-S Wang, "Deformation potentials at the top of valence bands in semiconductors: Ab initio pseudopotential calculations," *Phys. Rev. B* **46**, 12358 (1992).
- [12] J-Q Wang, B-Y Gu, "Intervalley $\Gamma - X$ deformation potentials for top valence bands in III-V zincblende semiconductors by ab initio pseudopotential calculations," *J. Phys. Condens. Matter* **5**, 647 (1993).
- [13] J. Harris, "Simplified method for calculating the energy of weakly interacting fragments," *Phys. Rev. B* **31**, 1770 (1985).
- [14] P. Vogl, "Microscopic theory of electron-phonon interaction in insulators or semiconductors," *Phys. Rev. B*, **13**, 694 (1976).
- [15] G. Gilat and Z. Kam, "High-Resolution Method for Calculating Spectra of Solids," *Phys. Rev. Lett.* **22**, 715 (1969).

A CPA Calculation for Disorder-Induced Intervalley Scattering in AlGaAs

Selim E. Günger and D. K. Ferry

Center for Solid State Electronics Research
Arizona State University, Tempe, AZ 85287-6206

Abstract

Disorder-induced scattering due to the random distribution of atomic species of the constituent semiconductors in alloys can lead to intervalley scattering transitions. A method based on a wavevector dependant formulation of the Coherent Potential Approximation is used to calculate the intervalley scattering potentials.

I Introduction

Recent experimental work of Kalt *et al.* [1] indicate that disorder scattering can have large wavevector components of the scattering potential, which may give rise to intervalley scattering. Grein *et al.* [2] also calculated the lifetime broadenings of electrons due to disorder-induced scattering from the Γ valley to the side valleys using an empirical pseudopotential formulation, and concluded that this is the dominant scattering process for electrons in the Γ valley minimum of AlGaAs with composition $x > 0.4$. In this study, we derive a general formulation for treating diagonal disorder-induced intervalley scattering transitions due to diagonal disorder based on the Coherent Potential Approximation (CPA). Previous formulations of the CPA for application to alloy bandstructures average out the self-energy over the entire Brillouin zone. In this work, the wavevector dependence of the relevant quantities are retained. The scattering potential $V_{sc}(q)$, which causes a scattering transition with wavevector q , is found by calculating the imaginary part of the CPA self-energy at the q point of interest.

II Coherent Potential Approximation

The coherent potential approximation method has been used previously to treat the electronic structure of alloys [3-5]. The exact Hamiltonian for an alloy is written as the sum of an ordered virtual crystal Hamiltonian, whose solution is known, and a random potential due to the random distribution of atoms:

$$H_{\text{alloy}} = H_0 + V_{\text{random}}, \quad (1)$$

$$V_{\text{random}} = \sum_{\mathbf{r}} (V_{\mathbf{r}} - V_0). \quad (2)$$

The random part of the crystal potential is replaced with an effective potential (self-energy) in the CPA approximation, which is described by an effective Hamiltonian

$$H_{\text{eff}} = H_0 + \Sigma(\mathbf{k}). \quad (3)$$

From the iterative solution of the Dyson's equation relating the Green's functions for the exact alloy Hamiltonian and the effective Hamiltonian

$$G_{\text{alloy}} = G_{\text{eff}} + G_{\text{eff}}(V_{\text{random}} - \Sigma)G_{\text{alloy}}, \quad (4)$$

the atomic scattering matrix, T , can be defined as

$$T = \frac{V - \Sigma}{I - (V - \Sigma)G_{\text{eff}}}, \quad (5)$$

which transforms (4) into the form

$$G_{\text{alloy}} = G_{\text{eff}} + G_{\text{eff}}TG_{\text{eff}}. \quad (6)$$

Taking the configuration average of both sides,

$$\langle G_{\text{alloy}} \rangle = G_{\text{eff}} + G_{\text{eff}}\langle T \rangle G_{\text{eff}}, \quad (7)$$

results in a self-consistency relation for the self-energy $\Sigma(\mathbf{k})$. This leads to the condition that the configuration average of the atomic scattering matrix vanish at every point in the Brillouin zone:

$$x \frac{V_A(\mathbf{k}) - \Sigma(\mathbf{k})}{I - (V_A(\mathbf{k}) - \Sigma(\mathbf{k}))G_{\text{eff}}(\mathbf{k})} + (1-x) \frac{V_B(\mathbf{k}) - \Sigma(\mathbf{k})}{I - (V_B(\mathbf{k}) - \Sigma(\mathbf{k}))G_{\text{eff}}(\mathbf{k})} = 0 \quad (8)$$

$V_A(\mathbf{k})$ and $V_B(\mathbf{k})$ are the anti-bonding potentials of the A and B semiconductors, calculated from the conduction band energy solutions, and are referenced with respect to the vacuum level. Chen *et al.* [3] use an approximation by which the effective Green's function, and the self-energy term are diagonalized. The effective Green's function is of the form of Bloch sums, which are valid for tight-binding, or bond-orbital basis functions and result in the cancellation of terms in the Brillouin zone integration. Then, the effective Green's function is of the form

$$G_{\text{eff}}(\mathbf{k}) = \sum_{\mathbf{k}'} \frac{1}{E_{\mathbf{k}} - E_{\mathbf{k}'} - \Sigma_{\mathbf{k}'}} \quad (9)$$

By discretizing the Brillouin zone into N points, we obtain an $N \times N$ system of coupled nonlinear equations for $\Sigma_{\mathbf{k}}$ to be solved self-consistently. Through a local \mathbf{k} approximation, we first calculate the density of states only at the \mathbf{k} -point of interest, through which the nonlinear system of equations can be decoupled. Rewriting the effective Green's function as

$$G_{\text{eff}}(\mathbf{k}) = -\frac{1}{\Sigma_{\mathbf{k}}} + \sum_{\mathbf{k}'}' \frac{1}{E_{\mathbf{k}} - E_{\mathbf{k}'} - \Sigma_{\mathbf{k}'}} \quad (10)$$

the summation term is neglected initially. We have checked that this approximation is valid, and find that the first term is about five times larger than the summation term. The resultant self energies $\Sigma_{\mathbf{k}}$, from the solution of (8) are fed back into the Green's function expression of (9) and the iteration is performed until self-consistency is reached. The self-energies do not change appreciably after the first iteration.

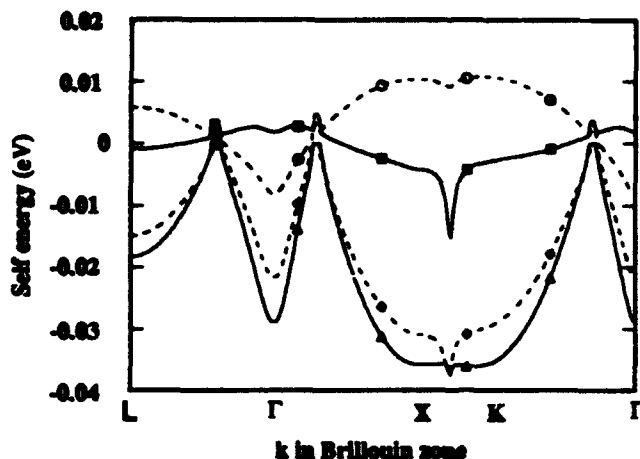


Figure 1. CPA Self-Energy in the Brillouin zone for $x = 0.2$ and 0.6 in AlGaAs. Circles and diamonds are the real and imaginary components of the self-energy for $x = 0.2$, and the squares and triangles are the real and imaginary self-energies for $x = 0.6$.

Using the sp^3s^* semi-empirical tight-binding method to calculate the eigenenergies, the scalar equations resulting from the above approximations are solved using a Newton-Raphson iteration for $\Sigma(k)$ at the discretized points in the irreducible $1/48$ -th wedge of the Brillouin zone. The results for the real and imaginary components of the self-energy are shown in Fig. 1 for two different compositions of AlGaAs. It is seen that the real part of the self-energy, which is a band-renormalization term in the virtual crystal approximation, results in a correction to the band energies whose magnitude is less than 10 meV throughout the Brillouin zone. The imaginary part is equal to the scattering potential, and shows a structure related to the symmetry points Γ , L and X . Between the X and K points, we have a dip in the self-energy, which is actually a result of the semi-empirical tight binding approximation to the bandstructure. In an sp^3s^* basis, the energy bandstructure between the X and K points has a degeneracy point, while in an Empirical Pseudopotential calculation of the band structure of AlGaAs, such a degeneracy point is not seen. The composition dependence of the imaginary part of the self-energy for the wavevector components L and X is shown in Fig. 2. There is a marked deviation from a $\sqrt{x(1-x)}$ type of behaviour, which is expected from a binomially distributed random atomic energy. Thus, the bowing parameter is not a simple scalar quantity.

III Applications

We apply our disorder-scattering potential results to the simulation of a femtosecond time-resolved pump probe experiment of intervalley scattering in $Al_{0.6}Ga_{0.4}As$ by Wang *et al.*[6]. The intervalley scattering transition is an L -wavevector transition, whose scattering potential is seen to be 18 meV from Figs. 1 and 2. We include this scattering process in an Ensemble Monte Carlo simulation of intervalley scattering in AlGaAs. The data points in Fig. 3 are the X -valley population in arbitrary units. A 400 femtosecond FWHM, 2.12 eV laser pulse excites carriers into the L and X valleys (the laser pulse is centered at $t = 1000$ femtoseconds in the

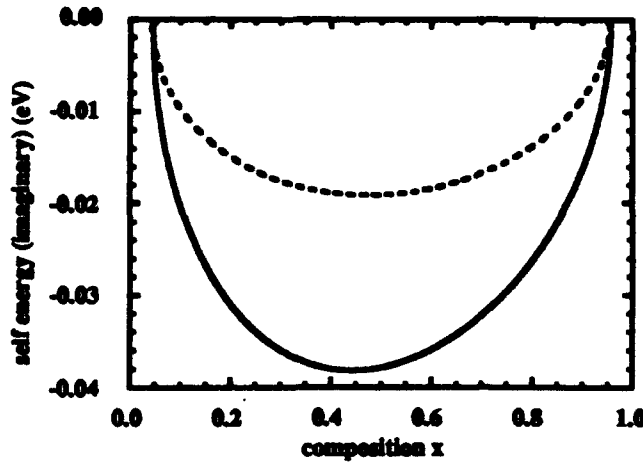


Figure 2. Imaginary part of the CPA self energy at the L point (dashed line) and the X point (solid line).

figure). Photo-excited carriers undergo inter- and intra-valley phonon- and disorder-induced scattering events to pile up at the bottom of the X valley. The population of the X valley is monitored by a 500 femtosecond FWHM convolving pulse which causes an infrared transition to the X_7 valley. Previously, we had reported a parameter fit of the non-polar optical deformation potential constant to be $D_{LX} = 1.5 \pm 0.5 \times 10^8$ eV/cm [7]. The results of the simulation for a value of $D_{LX} = 1.5 \times 10^8$ eV/cm per binary alloy mode are shown in Fig. 3 for three different disorder-induced intervalley scattering potentials. It can be seen that the effect of including disorder-induced scattering processes for L-X intervalley scattering is small.

IV Conclusion

Disorder-induced intervalley scattering is investigated using a Coherent Potential Approximation to model the random potential in the alloy AlGaAs. The imaginary part of the self-energy correction is responsible for scattering so we can use this self-energy to derive scattering potentials. In particular, a k -dependant formulation for the self-energy gives us the disorder potential components directly responsible for the large wavevector scattering transitions which can result in intervalley transfer. The disorder-induced intervalley scattering potential for Γ -X transitions (an X-wavevector) for the experimental case of [1] is found to be 38 meV at composition $x = 0.4$, somewhat smaller than the value reported (53 meV). However, this value seems overestimated from the derivation in [1]. Intervalley scattering between the L and X valleys has only a weak disorder-induced scattering potential. For AlGaAs with composition $x = 0.6$, disorder-induced scattering processes for L-X intervalley scattering are small compared with phonon-induced processes.

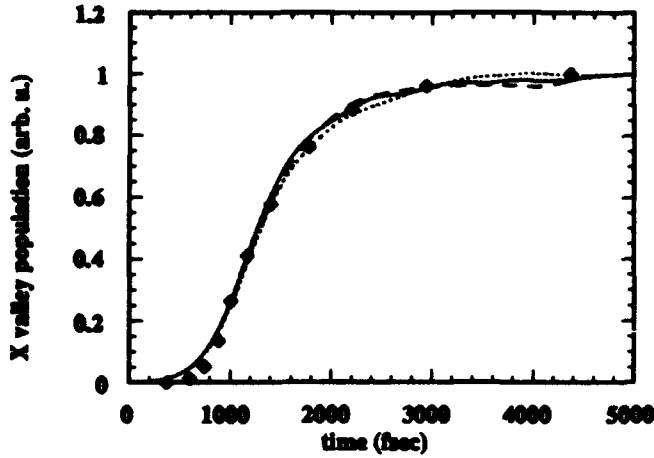


Figure 3. Effect of disorder-induced L-X intervalley scattering. The dotted line, solid line and the dashed line are for scattering potentials $V_{sc}(L) = 0, 18 \text{ meV}$ and 25 meV respectively. The diamonds are the data points of Wang *et al.* [6].

References

- [1] H. Kalt, W.W. Rühle, K. Reimann, M. Rinker, and E. Bauser, Phys. Rev. B **43**, 12364 (1991).
- [2] C. H. Grein, S. Zollner, and M. Cardona, Phys. Rev. B **44**, 12761 (1991).
- [3] A. B. Chen, and A. Sher, Phys. Rev. B **17**, 4726 (1978).
- [4] Y. T. Shen, and C. W. Myles, J. Phys. Chem. Solids **48**, 1173 (1987).
- [5] M. Bugajski, A. M. Kontkiewicz, and H. Mariette, Phys. Rev. B **28**, 7105 (1983).
- [6] W. B. Wang, K. Shum, R. R. Alfano, D. Szymd, and A. J. Nozik, Phys. Rev. Lett. **68**, 662 (1991).
- [7] S. E. Günçer, and D. K. Ferry, Phys. Rev. B **46**, 15309 (1992).

Steady-State Electron Transport in Silicon Dioxide Employing Different Electronic Band-Structures

M. Hackel, H. Kosina, and S. Selberherr

Institute for Microelectronics
Technical University of Vienna
Gusshausstrasse 27-29, A-1040 Vienna, Austria

Abstract

A semiclassical Monte Carlo technique is employed to simulate the steady-state electron transport in silicon dioxide at intermediate and high electric fields. The electronic structure is modelled by a single parabolic, by a single nonparabolic as well as an isotropic four-band model. We find that the electronic behavior of silicon dioxide is mainly influenced by a single nonparabolic conduction-band. The injection of electrons into silicon dioxide is also investigated in order to extract the thermalization length of electronic carriers.

I. Introduction

Silicon dioxide is of vital interest for "metal oxide semiconductor" (MOS) technology because of its importance as an insulator for gate electrodes. For "Ultra Large Scale Integrated" (ULSI) circuits the thickness of the insulating film is only comprised of a few nanometers, which results (for typical bias voltages) in a normal field approximately of 1 – 10 megavolts per centimeter (MV/cm). Experimental and theoretical studies give evidence that material breakdown will not occur under the influence of applied field-strengths as high as 20 MV/cm [1][2]. Under these enormous fields, the electronic distribution becomes unstable if the energy being gained from the field can no longer be given to the lattice. Three different scattering events are involved to model the transport behavior of electrons in SiO₂, namely, polar longitudinal optical (LO) phonons, nonpolar optical and nonpolar acoustic phonons [1][3][4][5][6][7][8].

II. The Physical Model

In low electron energy levels, polar longitudinal optical phonons are the dominant scattering process in silicon dioxide. The electrons lose a large amount of energy to the lattice due to a strong interaction of the polar phonon modes via the polarization field of the ions, but at high and intermediate fields they cannot prevent "velocity runaway" or even material destruction. Nonpolar acoustic phonons force electrons to scatter and stabilize the electronic distribution. As electrons reach the threshold for the emission of nonpolar acoustic phonons, the probability of having Bragg reflections (Umklapp-processes) increases keeping electronic carriers from gaining more energy from the field [3][7][9].

III. The Transport Model

In this section we briefly describe our Monte Carlo algorithm to solve the Boltzmann transport equation (BTE). An excellent review on this method is given in [10]. Some minor improvements and differences of this numerical method will be briefly reported here. The equation of motion and the duration of free flight are solved simultaneously by employing a Runge-Kutta algorithm instead of the usual self-scattering scheme. After performing a free flight the scattering process is randomly chosen according to the partial scattering rates. If one scattering process is selected the after-scattering state of the electron is calculated. For nonpolar electron-phonon collisions the polar θ and azimuthal angle ϕ are uniformly distributed [10], whereas the LO phonons favor large angle scattering. To compute the polar angle we use a modified rejection technique with a nonconstant enveloping function. The azimuthal angle of the LO phonons is chosen randomly. Having calculated the state of the scattered electron we perform another free flight till the maximum number of scattering events is reached.

IV. Results

To model the band-structure in SiO₂ we implemented an isotropic four-band model with one nonparabolic (nonparabolicity α) and three parabolic bands [11]

$$\epsilon(1 + \alpha\epsilon) = \frac{\hbar^2 k^2}{2m^*} \quad \text{for } 0 \leq k \leq k_{max} \quad \text{band 1,} \quad (1)$$

$$\epsilon = \epsilon_0 \pm \frac{\hbar^2 k^2}{2m^*} \quad \text{for } 0 \leq k \leq k_{max} \quad \text{band 2, 3, 4.} \quad (2)$$

band	$m^* [m_e]$	$\epsilon_1 [eV]$	$\epsilon_2 [eV]$	$k_{max} [nm^{-1}]$	multiplicity
1	0.50	0.00	5.52	11.54	6
2	1.34	5.52	9.31	11.54	6
3	1.05	7.00	9.00	7.42	12
4	1.05	9.00	11.00	7.42	12

Table 1. Parameters of the four-band model used

The parameters of the band-structure are summarized in Table 1. The data were extracted from band calculations of Chelikovsky and Schlüter [12].

In contrast to the electronic character of bands one and three, the bands two and four show hole-like behavior. The density of states (figure 1) of one parabolic, one nonparabolic and the four-band model are compared, whereas the first bands have the same mass. It is clearly seen for intermediate and high energies that the density of states strongly differs. The main features of a realistic band-structure with two maxima at 5.5 eV and 9 eV and one minimum at 7 eV are well reproduced and strongly influence the electronic distribution at all energies.

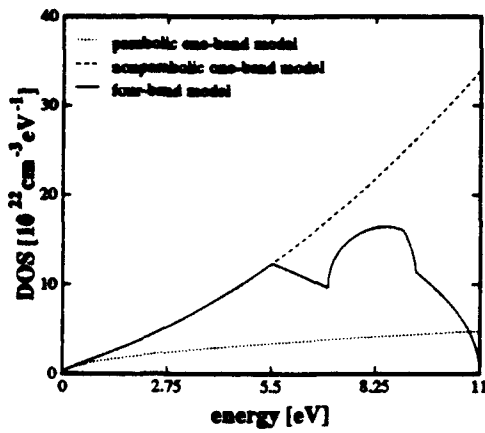


Figure 1. DOS of different band-structures used for SiO₂.

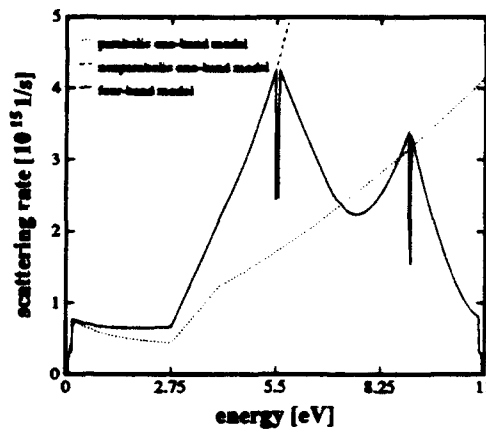


Figure 2. Scattering rates of different band-structures used for SiO₂.

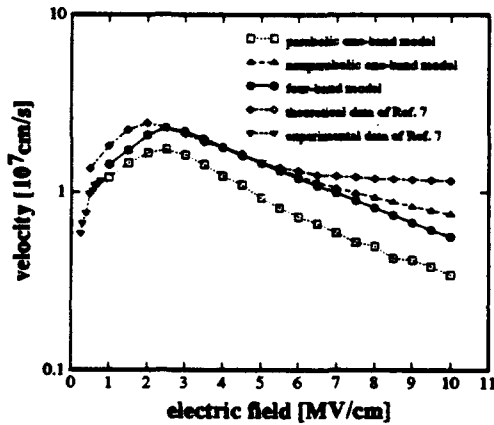


Figure 3. Drift velocity of different band-structures in SiO₂.

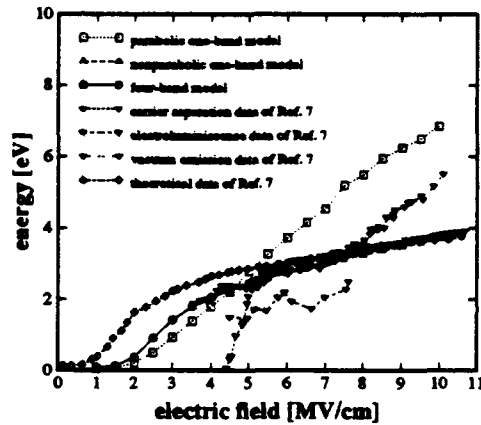


Figure 4. Mean kinetic energy in SiO₂ compared with the data of Ref. 7.

Figure 2 presents the total scattering rate for a temperature of 300 K and compares it with a single conduction-band. Nonpolar acoustic phonons set in at about 2.75 eV as the dominant scattering process (U-process). Compared with one-band models the different character of the four-band model again results in two maxima and one minimum. The discontinuity at the peak reflects the intravalley character of U-processes.

The dependence of the drift velocity of electronic carriers in SiO₂ versus electric field is plotted in figure 3. It increases till U-processes occur. A parabolic one-band model tends to lower velocities, whereas nonparabolicity increases the velocity. The split-off between the nonparabolic band-model and the four-band model is caused by a non-negligible occupancy of the second band at high electric fields. Our data are compared with the results of Fischetti [7][13].

The energy is plotted in figure 4. We observe that a single nonparabolic conduction-band and the four-band model do not exhibit any deviation, moreover, they almost demonstrate

quantitative identical values. Three different techniques have been employed to extract the energy as a function of the applied electric field, namely the carrier-separation technique, the electroluminescence method and finally the vacuum-emission technique [1],[2],[7].

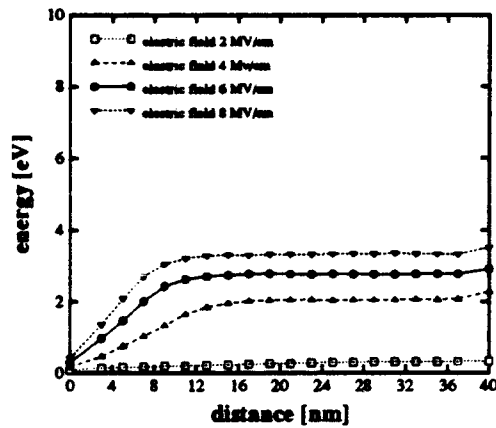


Figure 5. Spatial energy distribution in SiO_2 for electrons injected with 0.1 eV .

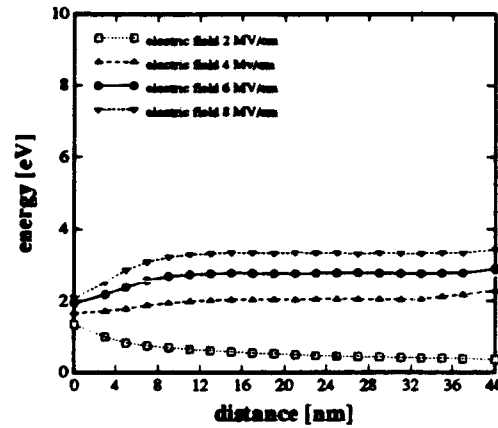


Figure 6. Spatial energy distribution in SiO_2 for electrons injected with 1 eV .

In figure 5 we investigate the thermalization length of electrons in SiO_2 employing one nonparabolic band. Electrons are injected at the left boundary according to an Boltzmann distribution with an average energy of 0.01 eV . We find that the thermalization length is dependent on the applied electric field as well as on the energy of the injected electrons. At low field-strengths the distribution shows that the electrons requires an average distance of about 30 nm to obtain the mean kinetic energy, whereas at high fields the average distance to thermalize is obviously lower than 10 nm . If the average energy of the applied field does not reach the threshold of U-processes the scattering rate is rather low and the mean free path large. Therefore scattering events are rarely resulting in long thermalization lengths. For high fields the mean free path is short and U-processes are dominant favoring large-angle scattering, which thermalize the carriers within few nm .

Figure 6 presents the distribution of electrons in SiO_2 that are injected with 1 eV in average. Again, we observe that the thermalization length of electrons is dependent on the electric field, but shorter than for carriers injected with low energy. The average length for electrons to thermalize is lower than 15 nm . For high fields it is clearly seen that electrons thermalize within 10 nm . In contrast to injected electrons with low energy the probability of suffering U-processes is high leading to a short mean free path and a short thermalization length.

Acknowledgement

The financial support of SIEMENS, Germany, is gratefully acknowledged. The authors are indebted to DIGITAL Equipment Corporation, USA, for providing an excellent computer environment.

References

- [1] M.V. Fischetti. *Physical Review Letters*, 53(18):1755-1758, 1984.
- [2] M.V. Fischetti and D.J. DiMaria. *Physical Review Letters*, 55(22):2475-2478, 1985.
- [3] B.K. Ridley. *Journal of Applied Physics*, 46(3):998-1007, 1975.
- [4] W. Porod and D.K. Ferry. *Physical Review Letters*, 54(11):1189-1191, 1985.
- [5] H. Köstner Jr. and K. Hübner. *Physica status solidi (b)*, 118:293-301, 1983.
- [6] H.-J. Fitting and J.-U. Friemann. *Physica status solidi (a)*, 69:349-358, 1985.
- [7] M.V. Fischetti, D.J. DiMaria, S.D. Brorson, T.N. Theis, and J.R. Kirtley. *Physical Review B*, 31(11):8124-8142, 1985.
- [8] M.V. Fischetti and D.J. DiMaria. *Solid-State Electronics*, 31(3/4):629-634, 1988.
- [9] M. Sparks, D.L. Mills, A.A. Maradudin, L.J. Sham, E. Loh Jr., and D.F. King. *Physical Review B*, 24(6):3519-3536, 1981.
- [10] C. Jacoboni and L. Reggiani. *Review Modern Physics*, 55(3):645-705, 1983.
- [11] R. Brunetti, C. Jacoboni, F. Venturi, E. Sangiorgi, and B. Ricco. *Solid-State Electronics*, 32(12):1663-1667, 1989.
- [12] J.R. Chelikowsky and M. Schlüter. *Physical Review B*, 15(8):4020-4029, 1977.
- [13] M.V. Fischetti, S.E. Laux, and D.J. DiMaria. *Applied Surface Science*, 39:578-596, 1989.

New Developments in Monte Carlo Device Simulation

Umberto Ravaioli

Beckman Institute
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA

Abstract

Monte Carlo particle methods have a unique role in semiconductor device simulation, since they allow one to solve the Boltzmann equation statistically, with the inclusion of many physical details which cannot yet be included completely in other approaches. Of the main limiting drawbacks of the technique, memory requirements and computational costs are much alleviated by the increasing power of computers. The problems due to statistical noise can in some applications be corrected by *ad hoc* techniques. This paper briefly reviews the state-of-the-art of Monte Carlo device simulation and elaborates on the future applications of the method.

I. Introduction

Monte Carlo particle simulation has been a powerful tool for the investigation of transport in semiconductors for well over twenty years. The evolution of Monte Carlo methods has been directly influenced by advances in computers, which have made possible the implementation of more and more refined physical models [1,2]. Originally, applications were limited to tracking the evolution of a single particle, to obtain steady-state time-averages of transport parameters. The single particle Monte Carlo technique is adequate to study bulk semiconductor properties under uniform field conditions, yielding as typical results the distribution function, average velocity and energy, valley occupation percentage, and velocity-field characteristics. The single particle model is also adequate to obtain local information in device structures for which a potential distribution is approximately known.

Transient and selfconsistent simulations were implemented when the memory of computer was increased to allow the simultaneous tracking of thousand of particles. The so-called ensemble Monte Carlo technique has then made possible a whole new range of selfconsistent applications which have required the inclusion of methods for the local evaluation of electronic forces, e.g. Poisson's equation [3]. The early simulation models have treated the bandstructure with an analytical parabolic or non-parabolic approximation. The energy range for hot electron analysis has been considerably extended with the introduction of algorithms which implement numerically a complete bandstructure of the semiconductor material [2,4]. Initially limited to bulk material and single particle applications, the full bandstructure Monte Carlo has been extended in recent years to ensemble selfconsistent applications and can be now run fairly efficiently on workstations [4,5]. While hardware improvements are making Monte Carlo applications more realistic, many efforts have been devoted to overcome the natural limitations of the technique, to optimize the algorithms and to take advantage of new hardware capabilities to introduce more advanced physical models.

The flow-chart of a selfconsistent Monte Carlo device simulation is relatively simple. The method uses a time-dependent approach, which besides providing transient results

may also be run until a steady-state is achieved. The iteration oscillates between a block which utilizes the information on charge density to evaluate the electrical forces in space, and a block which tracks the particle movement within a given timestep. The frequency of forces update is chosen as a trade-off between accuracy of the physics which requires very frequent force recalculation, and overall efficiency.

The particle movement is divided into two distinct parts: free flight under the influence of the electrical forces, and scattering events that terminate the flights. The length of the free flight trajectory is determined statistically, by relating the total scattering probability rate to a pseudo-random number picked from a uniform sequence generated by the computer. Once the flight is terminated, random number techniques are again used to select the type of scattering, according to the relative rate strength of the various mechanisms at that particular energy, and to determine the final state after the scattering event.

The particles are treated as classical objects obeying Newtonian mechanics during the free flights, and the scattering events are assumed to be instantaneous. Just a few lines of code are necessary to evaluate the momentum evolution during a timestep, using the classical law of accelerated motion. In an analytical band formulation, the energy at the end of the timestep is directly computed by evaluation of a simple formula. In the full bandstructure formulation, since the energy values are available only on a 3-D grid in the Brillouin zone of momentum space, the energy at the end of the timestep has to be evaluated by interpolation. This process is one of the major bottlenecks in the simulation, and innovative gridding approaches have been recently applied to implement faster energy evaluation techniques [6].

The weights of the different parts of the code, in terms of the overall CPU time, vary according to the implementation and running conditions. In most applications, the force evaluation by solving the Poisson's equation represents only a few percents of the computation, when it is solved at typical time intervals of 10 fs. However, in some cases (e.g. high doping) the time between two Poisson solutions must be reduced to 1 fs or less, to avoid spurious plasma oscillations of the particle gas, and in 2-D or 3-D simulations with a large number of grid points, such numerical solutions may carry a considerable weight. When forces are evaluated by a full molecular dynamics approach, the computational cost for this may be dominant, although there are multipole techniques which can be utilized to dramatically reduce the CPU time without sacrificing precision [7,8].

II. Statistical noise

Statistical noise due to the randomness of the events and discreteness of the particles always affects Monte Carlo simulation results. In many cases, it is sufficient to increase the number of particles and average in time the ensemble averages, to improve the accuracy of the collected statistics for various observables. However, there are cases where a mere increase of the number of samples is not practical. This is typically true in the case of effects which depend on high energy tails of the distribution function. An example is the injection of carriers in the oxide of a MOSFET structure. Steady-state parameters like potential and carrier distribution in the device are not very much affected by these rare events, but the evaluation of gate currents is exclusively determined by them. It is necessary to assign different weights to particles in different energy ranges in order to emphasize the statistics of high energy tails, while preserving the overall physical charge for force evaluation.

If the transient behavior is of interest, it is not possible to perform time averages and only an increase of the size of the ensemble would improve the statistics, but again,

this is seldom practical. Small-signal parameters of microwave devices can be determined for instance by Fourier transform of the transient currents. Deterministic models like drift-diffusion, which is not affected by statistical noise, are frequently used. They are also applied to study large-signal transient response of digital circuits. The Monte Carlo technique can be applied when drift-diffusion fails, provided that integrated quantities, rather than instantaneous ones, are considered. This is done by extending an idea first introduced by Hockney and Eastwood [2], where one keeps track of the total charge or current which transits through contacts. When the transient simulation is carried through steady-state, it is possible to identify two contributions to the cumulative charge, associated to the transient and to the steady-state regime [9]. Since the noise fluctuations are not very large for the time-integrated charge, it is possible to precisely fit in time the transient contribution with a polynomial or a combination of exponentials. This procedure yields very smooth curves for the transient currents, which can now be Fourier transformed to yield the intrinsic small-signal parameters.

III. Optimization

The core of the Monte Carlo algorithm determines the times of flight t by solving the integral equation $-\ln r = \int_0^t \lambda(t') dt'$, where r is a uniform random number between 0 and 1, and $\lambda(t)$ is the total scattering rate which changes in time as the particle momentum and energy vary under the influence of the fields. The choice of an appropriate solution is quite important for an efficient algorithm. An important concept is the *self-scattering*, a fictitious event which does not affect the electron state when selected. A self-scattering rate can be adjusted as convenient to facilitate the evaluation of the flight time, since it does not affect the statical properties of the process. The simplest approach is to add a self-scattering rate which makes the total rate constant, so that the integral above can be trivially determined.

A comparison of various techniques can be found in [10]. Self-scattering rates can be fixed as a function of energy, or can be dynamically adjusted in time along the particle trajectory. The goal of optimization of this process is to reduce as much as possible the amount of self-scattering while maintaining but still retaining it to guarantee that the solution of the integral is always statistically *correct*, rather than introducing approximations. The constant time technique [10,11] offers a good trade-off for self-consistent applications. The simulation time is divided into small constant steps for all the particles, and a self-scattering rate is adjusted to make the total rate constant in time within that interval. The integral is solved by adding up trivial contributions, until the equality is satisfied.

As mentioned earlier, the determination of the energy at the end of a timestep is a time consuming operation in full bandstructure calculations, which can create major bottlenecks. Optimization can be achieved by using a tetrahedral mesh in momentum space, instead of a regular grid, arranged so that nodes of a tetrahedron are positioned on adjacent energy isosurfaces [6]. With these grids, a linear expression can be used to determine directly the energy for given momentum coordinates, with precision controllable by adapting the grid locally. While the construction of this algorithm involves a considerable initial development cost, this technique promises to be a breakthrough which should make the full band approaches not more expensive than analytical band algorithms.

IV. Supercomputation

The advent of supercomputers has offered new opportunities to improve the performance of Monte Carlo codes and increase the size of the problems to be solved. The particle transport has an inherent parallel behavior which can be exploited, but the main obstacle is the fact that the histories of different particles can be very different due to the randomness of the processes, and such a difference can be quite emphasized in the case of sharply nonuniform structures.

Vectorization techniques can be very effective for bulk analysis, since it is sufficient to follow particles from one scattering event to another in parallel. The particle histories are equalized by sacrificing synchronism, which is not necessary for averaging. Synchronism must be maintained for self-consistent algorithms, because forces must be evaluated at specified intervals. Applications using an ensemble constant time approach for the flight time evaluation, yield a vectorization speed-up between 3-5 on a CRAY Y-MP supercomputer [11], where the maximum possible speed-up is about 10. Reports by several groups indicate similar speed-up for different Monte Carlo implementations. Comparisons in this area are extremely difficult, since when the efficiency of a code is improved, the achievable vectorization speed-up tends to decrease.

Parallelization can also be very advantageous for bulk calculations, since the particles are substantially independent. For more complicated self-consistent models, performance depends on the actual architecture of the hardware and on the strategy used to balance the load on the processors. Work in this area is still largely experimental, reflecting the immaturity of parallel computers and compilers. Parallelization is particularly appealing for 3-D simulation, because realistic applications require a very large number of particles [12]. Balancing of the load between processors is very important, because for massively parallel applications even a small percentage of non-parallelizable code may make the computation inefficient.

The future evolution of parallel architectures will have an important influence on Monte Carlo applications. This is particularly true for the most advanced applications which require the storage of large tables. The most memory intensive model is at the top of the hierarchy, where tables for the full bandstructure and momentum-dependent scattering rates must be stored and need to be equally accessible by all the processors. The information is only read by the processors during the simulation, since these tables are not changed. Therefore, for efficiency, such tables should reside in a shared memory region which can be quickly accessed by all nodes with uniform times, rather than being distributed in local memory areas appended to the processors or being completely copied in each of these areas, to limit storage requirements. The remaining compelling memory requirements are related to particle attributes, like position and momentum, which are continuously updated, and position dependent data (charge, fields) for large grids, particularly in 3-D. The main issue is to efficiently handle the communication between blocks of distributed memory. A logical storage scheme, in the case of distributed memory, is to map particles and grid nodes of domain subregions onto separate processors, trying to balance the number of particles per processor, adaptively throughout the simulation. As particles cross boundaries between subregions, they should be reassigned to new processors. For the determination of charge on grid nodes and the subsequent solution of Poisson equation, a small amount of communication is necessary between processors corresponding to physically contiguous regions. Much more challenging is the implementation of molecular dynamics, where in principle all processors need to communicate with each other. The optimal solution for a Monte Carlo code would be to have only shared memory available. In such a way it would be possible to load particles on fixed processors, regardless of position. Provided that each single processor can address a large shared memory necessary for the applications, still great challenges remain in designing system software and compilers for such a system.

V. Force Evaluation

The numerical solution of Poisson's equation only provides forces in a quasistatic approximation. In some applications this is not sufficient. It has been shown that to simulate fast phenomena associated with the transport of carriers generated by femtosecond laser pulses, the full Lorentz force must be evaluated [13], where the electric and magnetic fields should be obtained by solving the time dependent Maxwell's curl equations (note that we refer here the fields are generated by the fast moving charge particles, not to the laser radiation which is absorbed by the sample) Implementation of the algorithm can be very suitable to parallel computation and actually cheaper than solving the Poisson's equation for multi-dimensional simulations [12].

It is not clear at which frequency range the inclusion of the magnetic field begins to be necessary. In the simulation of general microwave devices, it should be possible to simply substitute Poisson's equation with the time-dependent wave equation for the retarded potential, which has the same space-dependent terms, to account for the displacement currents (in 1-D the displacement current contribution can be integrated and applied as additional boundary condition to Poisson's equation). However, in the THz regime the wavelength of the electromagnetic field in the doped semiconductor layers can be comparable or smaller than the dimensions of the active regions,

In other cases, the solution of Poisson's equation on discrete points may not be accurate enough to resolve the coulomb interaction between charges. Reduction of the mesh size may not improve much the situation, because the number of simulated carriers is fixed. A molecular dynamics can be used to calculate the force acting on a particle by adding the coulomb potential due to all the other charges. This approach would automatically include the electron-electron interaction effects, which can be very important in the case of high concentrations. The major computational obstacle is in the fact that the full molecular dynamics evaluation of the forces involves a number of operations of order N^2 , where N is the number of particles. In order to develop practical algorithms, it should be possible to apply multipole techniques, which have been extensively used to calculate molecule configurations, ionic systems [7] and capacitances in complicated VLSI interconnect layouts [8], to name a few applications. With accurate calibration, multipole algorithms only require a number of operations of order N . The idea is to consider particle-particle forces only within an appropriate neighborhood, and treat interactions from longer range particles through interpolated forces on a mesh with increasing coarseness at farther distances.

In many applications it is common to simulate only particles in a relatively small cell of a periodic structure, or a sample of a larger device. When the Molecular Dynamics method is implemented, the interactions with charges in other regions, which are not simulated, cannot be neglected. If one assumes that the simulated geometric sample is one element of a periodic structure, every particle in the simulated cell corresponds to a "replica" in each of the other cells of the periodic domain. The replicas of a given particle constitute then a "lattice" of charge, and special techniques must be applied to get a convergent sum of the interactions keeping down the computer time requirements [14]. For simulation of a bulk material, one has to consider 3-D lattices of replicas, 2-D and 1-D lattices for 1-D and 2-D device simulations, respectively, and in the case of a complete 3-D device simulation no replica has to be taken into account. Although more accurate than Poisson's equation, a molecular dynamics algorithm is still providing forces in an electrostatic approximation.

VI. Hybrid techniques

Monte Carlo techniques have found many useful applications beyond full self-consistent simulation of devices. Assuming that the *correct* solution of the Boltzmann equation is obtained, Monte Carlo simulations are often used to parameterize other models. Field dependent mobility and diffusion coefficient are used in drift-diffusion applications, for instance, and a number of other parameters are extracted to calibrate hydrodynamic and energy transport models. Although the Monte Carlo results should not always be trusted as completely exact, the potential problems in these schemes are more due to the fact that parameters, obtained for a bulk with uniform field, are often employed for nonuniform field condition, which can be considerably different.

The Monte Carlo approach is also used as a *postprocessor* using the potential profile obtained by a drift-diffusion or hydrodynamic approach. This is certainly a valid approach for large devices, where not only a drift-diffusion solution is acceptable, but also full self-consistent Monte Carlo solution would be impractical. By tracking particles simulated with Monte Carlo in the fixed potential, it is possible to evaluate high energy effects (injection into oxide, impact ionization) which are not well account for in simpler models. The advantage of a postprocessor is in the fact that one can use very efficient vector and parallel algorithms, since the tracked particles are essentially independent in nonself-consistent simulations.

A recent application of Monte Carlo simulation involves the calculation of tables for a scattering matrix technique [15]. The Monte Carlo procedures provides the possible outgoing momentum values, with associated probabilities, for a given incoming momentum into a thin slab of the device. By partitioning the device into slabs and matching the solutions at the interfaces with the momentum scattering tables, it is possible to get a solution which provides all the information of a self-consistent Monte Carlo, but with smooth solutions without the noise. Because of this, rare events should be easier to observe directly. Multi-dimensional applications are also possible with this technique.

VII. Improved Physical Models

A shortcoming of Monte Carlo models is due to the unavailability of many constants which are necessary to determine the scattering rates. A typical example is represented by deformation potentials of phonon scatterings. The usual procedure is to choose a set of deformation potentials which fit experimental data for steady-state velocity-field characteristic curves. Unfortunately, many slightly different sets can be found which provide a reasonable fit. Measurements of some important parameters, like the intervalley deformation potential for transitions between Γ and L valleys in GaAs, have been attempted, but the data reported by various groups are too contradictory to resolve the uncertainty.

The solution is to formulate new models which are based on first principles and rely less on parameter fitting. Electron-phonon scattering is usually treated with a simplistic dispersion relation, and is defined on arbitrary partitions of the Brillouin zone centered around energy minima (valleys). At high fields, the validity of this picture is questionable. A complete phonon model with accurate dispersion relations is needed, but the computational complexity is formidable. A consistent approach should treat both the bandstructure and the electron-phonon interaction on the same footage.

Electron-phonon matrix elements of the true Hamiltonian are equivalent to the matrix elements of a pseudo-Hamiltonian, as long as certain conditions on the true potential are met. Several calculations have been performed for semiconductors using local and nonlocal empirical pseudopotentials. These are assumed to be the sum of spherical potentials which

move rigidly with the atoms [16,17]. Another approach uses the *ab initio* pseudopotentials, which avoid self-consistency problems of the empirical approach by approximating the change in charge density when the atoms are displaced [18]. Computations using the Harris functional approach have provided the deformation potentials for phonon scattering in Si, throughout the Brillouin zone. A significant result is the evidence of the variation of deformation potentials with initial and final state wavevector. The total deformation potential exhibits a high degree of dispersion, especially for transitions away from the valley minima [18]. These results have significant implications for improving the predictive power of high-field Monte Carlo simulations.

VIII. Conclusions

Monte Carlo techniques for device simulation are undergoing dramatic developments due to the recent evolution of available computational platforms. Full bandstructure applications are already practical for use on top of the line workstations. New optimization techniques for the determination of momentum space trajectories, new approaches for electron-phonon interaction which remove much of the uncertainties of current models, and emerging applications on parallel architectures, should contribute to provide, in the next few years, accurate and efficient simulators with sufficient predictive capability in the high field transport regime to meet the needs of CAD designers of new generations of integrated devices. Monte Carlo techniques have already found a very important role as tools for the calibration of simpler models, and as postprocessing complements of conventional simulators, to quickly assess the importance of hot electron phenomena affecting device reliability.

Acknowledgments - This work has been supported by the Joint Services Program (grant N00014-90-J1270).

References

- [1] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation*, Springer-Verlag, Vienna, 1989.
- [2] K. Hess, ed., *Monte Carlo Device Simulation: Full Band and Beyond*, Kluwer Academic Publishers, 1991.
- [3] R.W. Hockney and J.W. Eastwood, *Computer Simulation using Particles*, McGraw Hill, New York, 1981.
- [4] M.V. Fischetti and S.E. Laux, "Monte Carlo Simulation of Transport in Technologically Significant Semiconductors of the Diamond and Zinc-Blende Structures-Part II: Submicrometer MOSFET's," *IEEE Trans. Elect. Devices*, vol. 38, pp. 650-660, 1991.
- [5] C.H. Lee and U. Ravaioli, "Monte Carlo Simulation of Si Devices", *Proc. VPAD '93*, pp. 36-39, Nara, Japan, May 14-15, 1993.
- [6] M.R. Pinto, J. Bude and C.S. Rafferty, "Simulation of ULSI Silicon MOSFETs", *Proc. VPAD '93*, pp. 22-25, Nara, Japan, May 14-15, 1993.

- [7] Z.A. Rycers, "Calculation of the Coulomb Interactions in Condensed Matter Simulation", *Molecular Simulation*, vol. 9, pp. 327-349, 1992.
- [8] K. Nabors, M. Kamon and J. White, "Multipole accelerated 3-D Interconnect Analysis", *Proc. VPAD '93*, pp. 72-75, Nara, Japan, May 14-15, 1993.
- [9] M.B. Patil and U. Ravaioli, "Transient Simulation of Semiconductor Devices Using the Monte-Carlo Method", *Solid-State Electronics*, vol. 34, pp. 1029-1034, 1991.
- [10] R.M. Yorston, "Free-Flight Time Generation in the Monte Carlo Simulation of Carrier Transport in Semiconductors," *J. of Comp. Phys.*, vol. 64, pp. 177-194, 1986.
- [11] U. Ravaioli, "Vectorization of Monte Carlo Algorithms for Semiconductor Simulation", Ch. 9 in *Monte Carlo Device Simulation: Full Band and Beyond*, K. Hess, ed., Kluwer Academic Publishers, 1991.
- [12] S. Pennathur, U.A. Ranawake, V.K. Tripathi, P. Lenders and S.M. Goodnick, "PMC-3D: A Parallelized 3D Monte Carlo Simulator for Electronic and Electro-optic Devices", these Proceedings.
- [13] K.M. Connolly, R.O. Grondin, R.P. Joshi, and S.M. El-Ghazaly, "Numerical Modeling of Ultrafast Electrical Waveform Generation and Characterization", *Proc. of SPIE Symposium on Ultrafast Laser Probe Phenomena in Bulk and Microstructure Semiconductors III*, vol. 1282, pp. 172-181, 1990.
- [14] D.K. Ferry, A.M. Krizan and M.J. Kann, "Molecular Dynamics Extensions of Monte Carlo Simulation in Semiconductor Device Modeling", *Computer Physics Communications*, vol. 67, pp. 119-134, 1991.
- [15] M.A. Alam, M.A. Stettler and M.S. Lundstrom, "Formulation of the Boltzmann Equation in Terms of Scattering Matrices", *Solid-State Electronics*, vol. 36, pp. 263-271, 1992.
- [16] M.V. Fischetti and J. Higman, "Theory and Calculation of the Deformation Potential Electron-Phonon Scattering Rates in Semiconductors, Ch. 5 in *Monte Carlo Device Simulation: Full Band and Beyond*, K. Hess, ed., Kluwer Academic Publishers, 1991.
- [17] T. Kunikiyo, Y. Kamakura, M. Yamaji, H. Mizuno, M. Takenaka, K. Taniguchi and C. Hamaguchi, "Adjustable Parameter Free Monte Carlo Simulation for Electron Transport in Silicon Including Full Band Structure", *Proc. VPAD '93*, pp. 42-43, Nara, Japan, May 14-15, 1993.
- [18] P.D. Yoder, V.D. Natoli and R.M. Martin, "*Ab Initio* Analysis of the Electron-Phonon Interaction in Silicon", *J. of Appl. Physics*, vol. 73, pp. 4378-4383, 1993.

Electron-Hole Scattering Effects in Silicon

Jim Dewey and Mohamed A. Osman

School of Electrical Engineering and Computer Science
Pullman, Washington, 99164-2752

Abstract

The role of electron-hole scattering on the low field mobility of electrons in *p*-doped silicon is examined using an Ensemble Monte Carlo approach. The transport model includes both electron and hole dynamics with coupling between the electron and hole systems through coulombic electron-hole interactions. The electron-hole scattering rate is calculated using the elliptical nature of the conduction band valleys and non-parabolicity of the valence band. Minority electron low-field drift mobilities are calculated at acceptor concentrations of $4.5 \times 10^{16} \text{cm}^{-3}$ and $3.8 \times 10^{18} \text{cm}^{-3}$ at 300K. Calculations are in excellent agreement with experimental values.

I. Introduction

Experimental investigations of minority electron transport in silicon report a considerable reduction in electron mobility in the presence of an applied electric field[1,2]. This decrease is attributed to the drag effect of electron-hole scattering which is a result of the net momentum transfer from electrons to holes. In this work, Ensemble Monte Carlo techniques are used to examine the low-field mobility of minority electrons in silicon.

Previous microscopic studies of minority electron transport using Monte Carlo techniques have been conducted for GaAs[3,4] and silicon[5,6]. Furuto and coworkers[3] compared experimental drift velocity in *p*-doped GaAs with Monte Carlo calculations. The hole transport model used in their work contained only heavy holes and simulated hole dynamics with a drifted Maxwellian distribution. Their hole model ignores the complex nature of the valence band. Sedra and coworkers[4] used a more accurate hole transport model and performed simulations using a single particle Monte Carlo technique. A shortcoming of both their works and our previous work in silicon[5,6] is the use of the Brooks-Herring impurity scattering model. This model is based on the first Born approximation and is invalid for impurity concentrations above 10^{17}cm^{-3} at room temperature. This was clearly shown in our previous calculations at $3.8 \times 10^{18} \text{cm}^{-3}$ [6] where calculated low field mobility overestimated experimental values[1]. A new impurity scattering model for Monte Carlo calculations was developed, to correct this problem[7].

In this contribution we report low field drift mobility of electrons in room temperature silicon for acceptor concentrations of $4.5 \times 10^{16} \text{cm}^{-3}$ and $3.8 \times 10^{18} \text{cm}^{-3}$ at 300K. Calculations are reported using an improved hole transport model[8] and impurity scattering model[7].

II. Transport Model

The transport model includes the dynamics of both electrons and holes. Modeling of both carriers is required, due to the strong effect of the hole plasma on electron transport. The hole transport model includes warped non-parabolic heavy and light hole bands and a spherical spin-orbit band. The non-parabolic nature of the heavy and light hole bands is represented by an energy dependent effective mass. The scattering mechanisms taken into account are intra and inter-band elastic acoustic and non-polar optical phonon scattering. More details on the hole transport model along with comparisons to experimental values are reported elsewhere[8].

The electron transport model contains all six non-parabolic ellipsoidal α -valleys. Three intervalley g -type phonon scattering between parallel valleys, three intervalley f -type phonon scattering between perpendicular valleys, elastic intervalley acoustic phonon scattering, ionized impurity scattering, and electron-hole scattering. The phonon scattering and material parameters given by Brunetti and coworkers[9] and the intra and inter-valley phonon scattering rates calculated by Jacoboni and Reggiani[10] are used in this work. Since calculations are reported at low fields and doping concentrations greater than 10^{17}cm^{-3} , impurity scattering models based on the first Born approximations, such as the Brooks-Herring model, overestimate experimental mobility. For this reason a new impurity scattering model was developed and implemented in our Monte Carlo program[7]. The model and method used to implement electron-hole scattering is discussed in the next section.

III. Electron-Hole Scattering Rate

The interaction between an electron and hole is assumed to be a screened coulomb potential. The probability of an electron and hole making a transition from the initial state $(\mathbf{k}_e, \mathbf{k}_h)$ to a final state $(\mathbf{k}'_e, \mathbf{k}'_h)$ as a result of a Coulombic interaction is given by the Fermi Golden rule for first-order transitions. By summing over all final electron and hole states and determining whether or not the final states are occupied after the scattering mechanism is chosen[10], the scattering rate becomes

$$S_{eh}(\mathbf{k}_e, \mathbf{k}_h) = \frac{Q^4 m_0 p}{2\pi \hbar^3 \epsilon^2 \beta^{*3}} \left(\frac{m_0}{\mu_d} \right)^{1/2} \int \frac{g^*}{(\beta^{*2} + g^{*2})} f_{\mathbf{k}_h} d\mathbf{k}_h \quad (1)$$

where m_0 is the free electron mass, p is the hole concentration, μ_d is defined as

$$\mu_d = (\mu_x \mu_y \mu_z)^{1/3} \quad (2)$$

where

$$\frac{1}{\mu_i} = \frac{1}{m_h} + \frac{1}{m_i} \quad (3)$$

and (i specifies the direction x, y, z)

$$g^* = (g_x^{*2} + g_y^{*2} + g_z^{*2})^{1/2} \quad (4)$$

where

$$g_i^* = (m_0 \mu_i)^{1/2} \left(\frac{\mathbf{k}_{e1}}{m_i} - \frac{\mathbf{k}_{h1}}{m_h} \right). \quad (5)$$

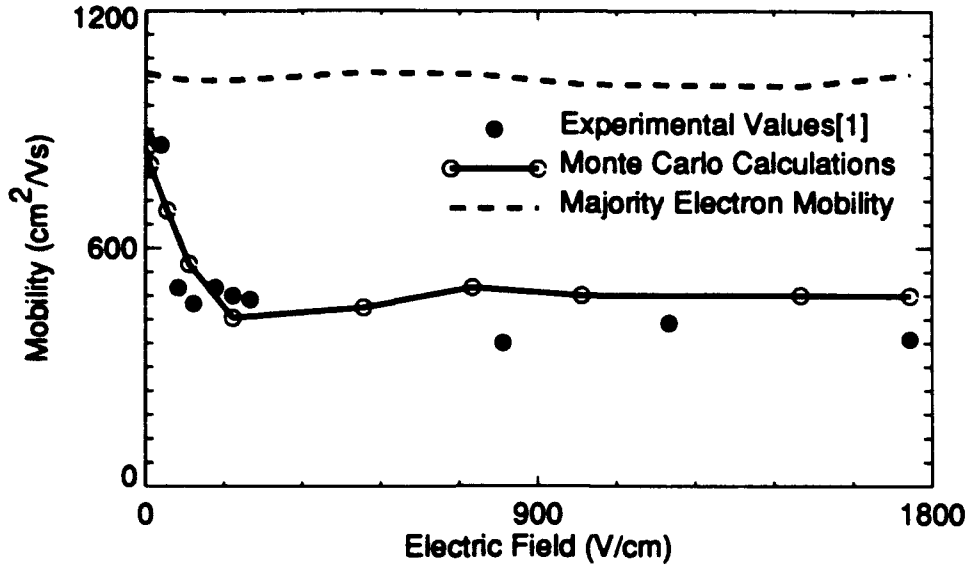


Figure 1: Minority Electron mobility in room temperature silicon as a function of applied electric field for an acceptor concentration of $4.5 \times 10^{16} \text{ cm}^{-3}$.

The relative wavevector g is made more complex by the direction dependent electron mass and energy dependent hole mass. In its present form, equation (1) is not suitable for calculating scattering rates in an Ensemble Monte Carlo program, since the hole distribution function is not known before the scattering event is chosen. To overcome this problem, the method proposed by Brunetti and coworkers[9] is used where the term $g^*/(g^{*2} + \beta^{*2})$ is replaced by its maximum value of $1/2\beta^*$. By using the maximum value, the resulting scattering rate is the scattering rate for electron-hole scattering plus a self-scattering rate and a rejection method is used.

IV. Results

Figures (1) and (2) show minority electron mobility calculations and experimental data as a function of applied electric field for acceptor concentrations of $4.5 \times 10^{16} \text{ cm}^{-3}$ and $3.8 \times 10^{18} \text{ cm}^{-3}$, respectively. Majority electron mobility is shown for comparison. As these figures show, there is good agreement between experimental results report by Tang and coworkers[4] and Monte Carlo calculations. The experimental results of Tang and coworkers[4] show a 45% decrease from the zero field mobility at only 100V/cm at both acceptor concentrations while Monte Carlo calculations show a 40% decrease under the same conditions. This dramatic decrease in mobility when an electric field is applied is due to the drag effect of electron-hole scattering. After the initial decrease, mobility essentially stays constant as the electric field is increased. This saturation of the drag effect is due to the inherent Coulombic nature of electron-hole scattering. As the electric field is increased, the velocity of both carriers also increase resulting in a larger exchange of momentum during electron-hole scattering. But, as with any Coulombic interaction, as the velocity increases the scattering cross section decreases. The larger exchange of

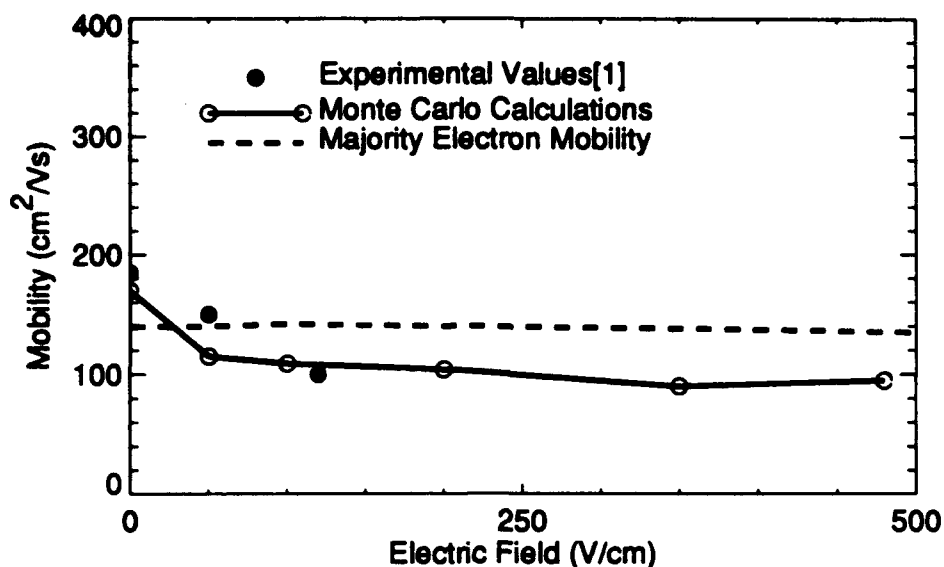


Figure 2: Minority Electron mobility in room temperature silicon as a function of applied electric field for an acceptor concentration of $3.8 \times 10^{18} \text{ cm}^{-3}$.

momentum is countered by a decrease in the frequency of the interactions resulting in a saturation of the drag effect.

V. Conclusion

Minority electron transport in room temperature silicon has been examined using an Ensemble Monte Carlo approach that includes the dynamics of both electrons and holes. A theoretical expression for electron-hole scattering which takes into account the ellipsoidal nature of the conduction band valleys and non-parabolicity of the valence band was developed and implemented. Improved hole transport and impurity scattering models were also implemented. Using the model discussed in sections (II) and (III), calculations are in excellent agreement with experimental results. Both calculated and experimental data show a dramatic decrease from the zero field mobility when the electric field is only 100V/cm. This reduction is attributed to the drag effect of electron-hole scattering.

VI. References

- [1] D.D.Tang, F.F.Fang, M.Scheuermann, and T.C.Chen, "Time-of-flight Measurements of Minority-Carrier Transport in *p*-silicon," Appl. Phys. Lett., vol. 49, pp.1540-1541, 1986.
- [2] M.Morohasi, N.Sawaki, and I.Akasaki, "Electron Mobility and Drag Effect in *p*-doped Silicon," Japan. J. Appl. Phys., vol. 24, pp. 732-736, 1985.
- [3] T.Furuta, H.Taniyama, M.Tomizawa, and A.Yoshii, "Hot-Carrier Transport in *p*

- GaAs," *Semicond. Sci. Technol.*, vol. 7, pp.B346-B350, 1992.
- [4] K.Sedra, C.M.Masiar, B.G.Streetman, and D.S.Tang, "Effects of Multiband Electron-Hole Scattering and Hole Wavefunction symmetry on Minority-Electron Transport in GaAs," *unpublished manuscript*.
 - [5] J.Dewey and M.A.Osman, "Monte Carlo Investigation of Minority Electron Transport in Silicon," *Proc. IWCE*, pp.225-228, 1992.
 - [6] J.Dewey and M.A.Osman, "Monte Carlo Investigation of Minority Electron Transport in Silicon," *Appl. Phys. Lett.*, vol. 62, pp.187-189, 1993.
 - [7] J.Dewey and M.A.Osman, "Impurity Scattering Model for Monte Carlo Calculations," (in preparation).
 - [8] J.Dewey and M.A.Osman, "Monte Carlo Simulation of Hole Transport in Silicon," *J. Appl. Phys.* (in press).
 - [9] R.Brunetti, C.Jacoboni, F.Nava, L.Reggiani, "Diffusion Coefficient of Electron in Silicon," *J. Appl. Phys.*, vol. 52, pp. 6713-6722, 1981.
 - [10] C.Jacobini and L.Reggiani, "Monte Carlo Method in Transport," *Rev. Mod. Phys.* vol. 55, pp.645-705, 1983.
 - [11] P.Lugli and D.K.Ferry, *Physica*, vol. 117B, pp.251-254, 1983.

Hole Transport in the Warped Band Model of GaAs

N. Nintunze and M.A. Osman

School of Electrical Engineering and Computer Science
Washington State University, Pullman, WA 99163-2752

Abstract

Hole transport in the warped band model of GaAs has been investigated using an Ensemble Monte Carlo approach. Scattering rates used in the simulations were derived for warped heavy and light hole bands and for a spherical split-off band. The obtained velocity-field characteristics are in better agreement with experimental data than recently published values where only the heavy and light hole bands were included in the model and warping was accounted for by an approximate overlap function. Simulations show that a better fit to the experiment and to results of a realistic band structure model can be attained with an optimal choice of valence band parameters.

I. Introduction

The performance of semiconductor devices, such as bipolar transistors, p-channel field-effect transistors and heterojunction bipolar transistors, is determined by hole transport properties. In some earlier studies, warping of the valence band has been either ignored[1] or taken into account by the use of an approximate overlap function[2]. The split-off band has also been mostly ignored due to its relatively low population. However, warping and the effect of the spin-orbit interaction on the heavy and light hole band cannot be overlooked if precise hole transport modeling is an issue. Brennan and coworkers[3, 4] used the Monte Carlo method in their investigation of hole transport in a realistic band structure. However, high precision is achieved that way at the expense of CPU time. Recently Brudevoll and coworkers[5] pointed to the fact that scattering rates often used in literature for ionized impurity scattering, inter-band polar optical phonon scattering and acoustic phonon scattering rates require corrections. Brudevoll and coworkers[2] studied the warm and hot hole drift velocity in GaAs using a valence band model made of a heavy and a light hole band, and warping of the valence bands was taken into account by the use of an approximate overlap function.

In this paper, hole transport in a warped valence band model of GaAs is investigated using an Ensemble Monte Carlo program. The scattering rate calculations for intra- and inter-band hole scattering take warping and the overlap integrals into account. Hole velocities were obtained for different sets of warping parameters and for fields up to 100kV/cm. Simulation results using this simple model are compared to the results of Brudevoll and coworkers[5], Brennan and coworkers[3], and to the experimental data of Holway and coworkers[6]. The Monte Carlo model used in this work is presented next. The results of the simulation are discussed in section III. Conclusions are given in section IV.

II. The Monte Carlo Model

The Monte Carlo approach is used in this work to investigate hole transport for fields applied along the (100) direction. The valence band model includes warped heavy and light hole bands and a parabolic split-off band. For the warped bands, the energy is given by:

$$E(k) = \frac{|A|\hbar^2 k^2}{2m_0} [1 \pm g(\theta, \phi)] \quad (1)$$

where

$$g(\theta, \phi) = [(B/A)^2 + (C/A)^2(\sin^2 \theta \cos^2 \theta + \sin^4 \theta \cos^2 \phi \sin^2 \theta)]^{1/2} \quad (2)$$

Above, the $+$ ($-$) sign correspond to the light(heavy) hole band, $\theta(\phi)$ is the polar(azimuthal) angle of k , and A , B , and C are the inverse band mass parameters. The set of parameters ($A=-6.98$, $B=-4.5$, $C=6.2$) given by Madelung[7], and the sets ($A=-7.98$, $B=-5.16$, $C=6.56$) and ($A=-7.65$, $B=-4.82$, $C=7.7$) used by Brudevoll and coworkers[5] are adopted in simulations. The mass of the holes in the warped bands is direction dependent and is given by:

$$m_h = \frac{m_0}{A(1 \pm g(\theta, \phi))} \quad (3)$$

For the split-off band, a spherical equi-potential surface is assumed with a scalar effective mass. Scattering mechanisms included are the acoustic phonon, optical phonon scattering, screened polar optical phonon, with intra- and inter-band transfers for all three bands. Impurity scattering is not included for simplicity.

The scattering rates contain overlap factors which are approximated by functions derived by Wiley[8]. The resulting expressions become easy to use in Monte Carlo calculations. In the particular case of ionized impurity scattering, Jacoboni and Reggiani[9] propose to neglect a weak warping in the process of scattering process determination, but use a corrected effective mass. Warping would be taken into account in final state determination. Such approach is justified as warping has a noticeable effect on next state determination rather than on the values of the scattering rates. Tiersten[10], for example, found that the valence band anisotropy can result in hole mobility variation of up to 20%, while the scattering rates vary by less than 1%. In our Monte Carlo program, scattering rates are evaluated at each time step so that angular evolutions could be observed. The scattering rates for the acoustic and optical non-polar scattering are weighted by an overlap factor of 1/2 while the polar optical transition rates are integrated taking the overlap factor into account. The final states are determined by applying the rejection technique to the exact angular probability functions containing angular contributions from both warping and the overlap integral. The results of the simulations will be presented in the following section.

III. Results and Discussion

The velocity-field characteristics were calculated for the current model at 300K lattice temperature. Figure 1 shows the obtained velocities for the same parameters used by

Brudevoll and coworkers[2]. Also shown are the experimental results of Holway and coworkers[6]. The recent results of Brudevoll and coworkers overestimate the hole drift velocity at all fields. It is obvious from figure 1 that the current calculations, which take into account the direction dependence of the hole mass, are in better agreement with the experiment. The inclusion of the split-off band doesn't have a strong effect at low fields as its hole population is low. However, its contribution to the hole velocity becomes important at fields higher than 60kV/cm. It is also above that field that the current model and the model of Brudevoll and coworkers don't realize a complete velocity saturation. Figure 2 illustrates the dependence of the hole drift velocity on warping parameters. The shown simulation results were obtained using 3 sets of valence band parameters and GaAs parameters taken from reference [4]. The set ($A=-6.98$, $B=-4.5$, $C=6.2$) by Madelung[7] and ($A=-7.65$, $B=-4.82$, $C=7.7$) result in almost equal velocity curves and band warping. These results overestimate the velocity at high fields when compared to experimental data and the results of Brennan and coworkers[3] obtained using a more complete valence band model. However, for fields below 60kV/cm results closer to the experiment can be obtained by an optimal choice of valence band parameters.

IV. Conclusion

We have investigated hole transport in the warped band model of GaAs, using an Ensemble Monte Carlo program. The velocity-field characteristics for three recently reported valence band parameters have been shown to be between the recent results of Brudevoll and coworkers[5], the results of Brennan and coworkers[4] and the experimental data of Holway and coworkers[6]. The current model, though simple, coupled to an optimal choice of valence band parameters, results in velocity-field characteristics consistent with previously published data. This model could be useful in the study of anisotropy dependent phenomena, such as orientational relaxation of carrier momentum.

References

- [1] M. Costato, C. Jacoboni and L. Reggiani, Phys. Stat. Sol.(b) 52, 461(1972)
- [2] T. Brudevoll, B. Lund, and T.A. Fjeldly, J. Appl. Phys. 71, 4972 (1992)
- [3] K. Brennan, K. Hess and Iafrate, J.Appl.Phys.55(10),3632(1984)
- [4] K. Brennan and K. Hess, Phys.Rev. B 29, No.10, 5581(1984)
- [5] T. Brudevoll, T.A. Fjeldly, J. Baek, and M. Shur, J. Appl. Phys. 67, 7373 (1990)
- [6] L.H. Holway, S.R. Steele and M.G. Alderstein, Proceedings of the Seventh Biennial Cornell Electrical Engineering conference(Cornell University Press, Ithaca, NY, 1979), pp.199-208.
- [7] O. Madelung(ed.), Semiconductors: Group IV Elements and III-V Compounds, Springer-Verlag, 1991

[8] J.D.Wiley, Phys. Rev. 4 No.8 2485(1971)

[9] C. Jacoboni and L. Reggiani, Reviews of Modern Phys. Vol.55,No.3, July 1983

[10] M. Tiersten, J.Phys.Chem.Solids, 25, pp.1151-1168, 1964

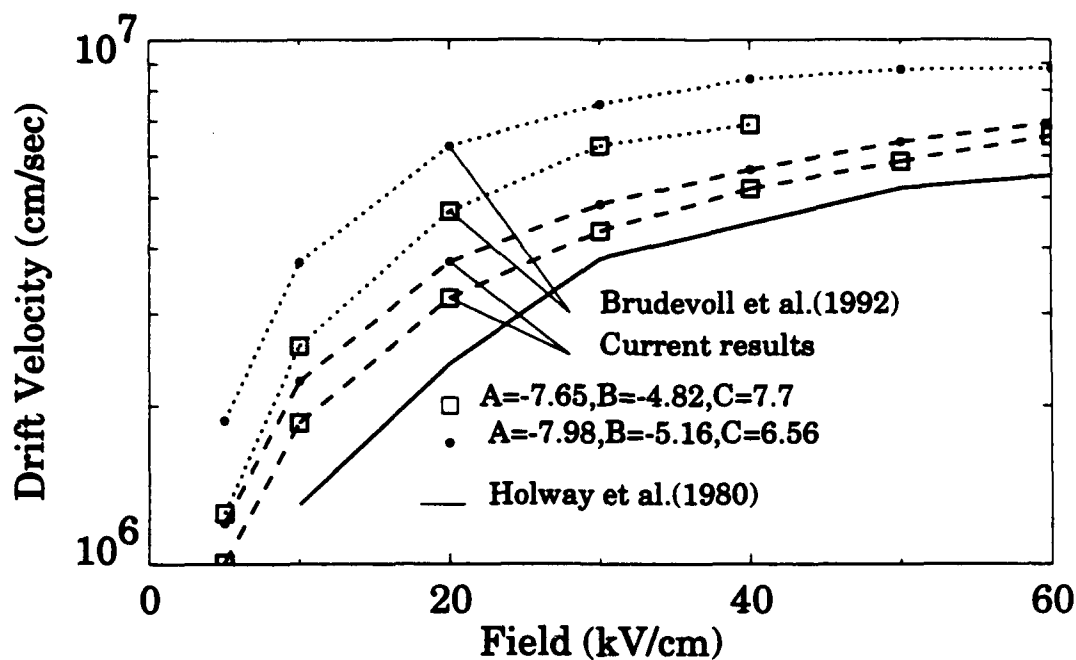


Fig.1. Velocity-Field characteristics for holes

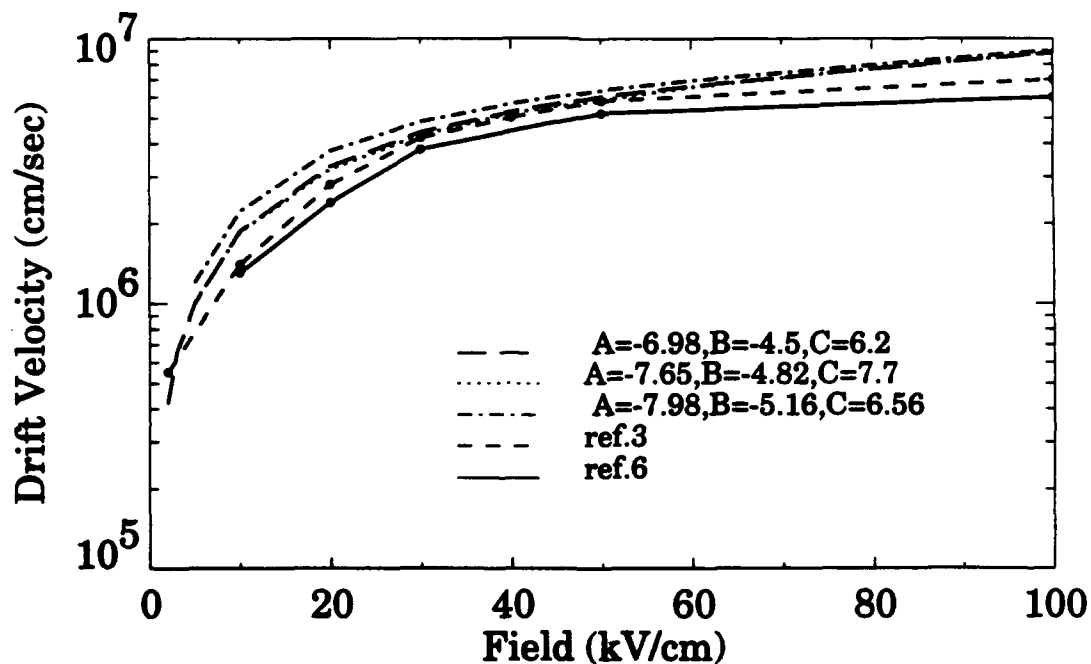


Fig.2. Velocity-Field characteristics for holes for different valence band parameters.

Ensemble Monte Carlo Simulation of the Hot Electron Transport in the Heterojunction Bipolar Transistors

Gregory Khrenov, Victor Ryzhii and Sergey Kartashov¹

Department of Computer Hardware,
The University of Aizu,
Aizu-Wakamatsu City, Japan

Abstract

The effect of the electron transport regime in the base-collector junction on the steady-state and high-frequency characteristics of a heterojunction bipolar transistor (HBT) is examined. A new technique based on the Fourier analysis of the induced collector current to evaluate HBT high frequency performance is used. The non-stationary electron transport, velocity overshoot effects and effects of the high-current density are taken into account using self-consistent time-dependent ensemble Monte Carlo particle simulation.

I. Introduction

Submicrometer heterojunction bipolar transistors (HBTs) are now widely investigated because of their great promise for microwave, millimeterwave and ultra-high-speed digital operations. In a typical $N-p-n$ HBT with abrupt emitter-base heterojunction electrons are injected from wide-gap emitter with considerable excess energy into a p -type narrow-gap base region. As a result the electron transport in the submicrometer HBT is far from equilibrium. To improve the HBT high-frequency performance the different base-collector junction designs have been proposed and investigated experimentally in [1,2] and using numerical simulation in [3-5]. For these designs the electrons stay longer in central valley and therefore achieve high average velocities into the collector region. Thus, extreme nonequilibrium electrons transport in the base and collector regions plays a role in determining device performances. Therefore to calculate transistor electrical characteristics with high accuracy the numerical simulation seems to be the only powerful way.

II. Numerical Model

To investigate the operation and design principles of the HBTs one-dimensional ensemble Monte Carlo simulator was developed [5]. Our Monte Carlo model incorporates complicated non-parabolic electron energy spectra and all essential scattering mechanisms. Motion of the holes in the base region and the electrons in the heavily doped collector contact region is evaluated using the drift-diffusion approach. Two distinctive features are inherent in the model. First feature is concerned with the choice of the self-scattering events. To minimize the number of the self-scattering events a special procedure is used where the total scattering rates are represented by the tabulated momentum-dependent step-like

¹Institute of Physics and Technology, Moscow, Russia

functions. These functions are chosen taking into account local values of the electric field and impurity concentration. Applications of this technique permits to reduce the number of self-scattering events to 7% of the scattering events total number.

Other novice of the model is concerned with the calculation of the HBT small-signal high-frequency performance. The HBT high-frequency performance is determined by the base resistance, the emitter and collector capacitances effects and the carrier transit effects in the base and collector regions. To characterize the HBT as an microwave and millimeter-wave amplifier the cut-off frequency f_T is usually used. As a rule the following expression derived from the small-signal equivalent circuit analysis is used to estimate f_T in a HBT [1]:

$$f_T = \{2\pi \cdot [\tau_E \cdot C_{BE} + \tau_{BC} + (\tau_E + R_E + R_C) \cdot C_{BC}]\}^{-1}, \quad (1)$$

where τ_E is the emitter resistance, R_E is the emitter series resistance, R_C is the collector series resistance, C_{BE} and C_{BC} are the emitter and collector capacitance respectively, τ_{BC} is the base-to-collector transit delay time. In this expression all the carrier transit effects are considered by the only carrier transit delay time through the base and collector τ_{BC} usually being estimated from the steady-state distribution of the mean electron velocity [1,6]. It is obvious that this τ_{BC} formulation is not absolutely correct because it does not take into account the electron velocity profile in the collector region and detail shape of the electron distribution function. The last reason is especially important in the presence of the non-stationary carrier transport. Therefore method based on the Fourier analysis of the non-steady-state induced collector current is used in present model.

Simple high-frequency equivalent circuit of HBT is shown in Fig.1. For this circuit frequency-dependent common-emitter short-circuit current gain h_{21} may be written as follows:

$$h_{21}(\omega) = \left[\frac{(1 + j\omega C_{BE}\tau_E) \cdot (1 + j\omega C_{BC}R_C)}{\gamma(\omega) - j\omega C_{BC}(\tau_E + R_E) + \omega^2 C_{BE}C_{BC}\tau_E R_E} - 1 \right]^{-1}, \quad (2)$$

$$\gamma(\omega) = \gamma_E(\omega) \cdot \beta(\omega) \cdot \gamma_C(\omega)$$

where γ_E is the emitter efficiency, β is the base transmission coefficient and γ_C is the collector transmission coefficient. The values of the resistances and capacitances in (2) are calculated from the numerical simulation of the required steady state or are estimated from the experimental results. Further in our analysis, C_{BC} , C_{BE} , τ_E , R_E , R_C and γ_E are assumed to be independent on frequency and to be dependent on the applied bias. The only problem is the calculation of the complex frequency-dependent coefficient $\gamma(\omega)$.

To calculate $\beta(\omega) \cdot \gamma_C(\omega)$ the Fourier analysis of the non-steady-state induced collector current is used [3]. Initially the steady-state simulation for the given applied bias is performed and stationary distributions of the potential and carriers concentrations are calculated. After that the electron bunch consists of several thousand electrons is launched from the base-emitter interface during short time interval and the time dependent collector current response $J_C(t)$ is calculated. Here we assume that the injected non-steady-state electrons travel through base and collector in the known stationary electric field. When all launched electrons leave the modeling region the Fourier transformation of the injected non-stationary emitter current $J_E(t)$ and induced collector current are carried out. The frequency-dependent base-collector transmission coefficient is defined as follows:

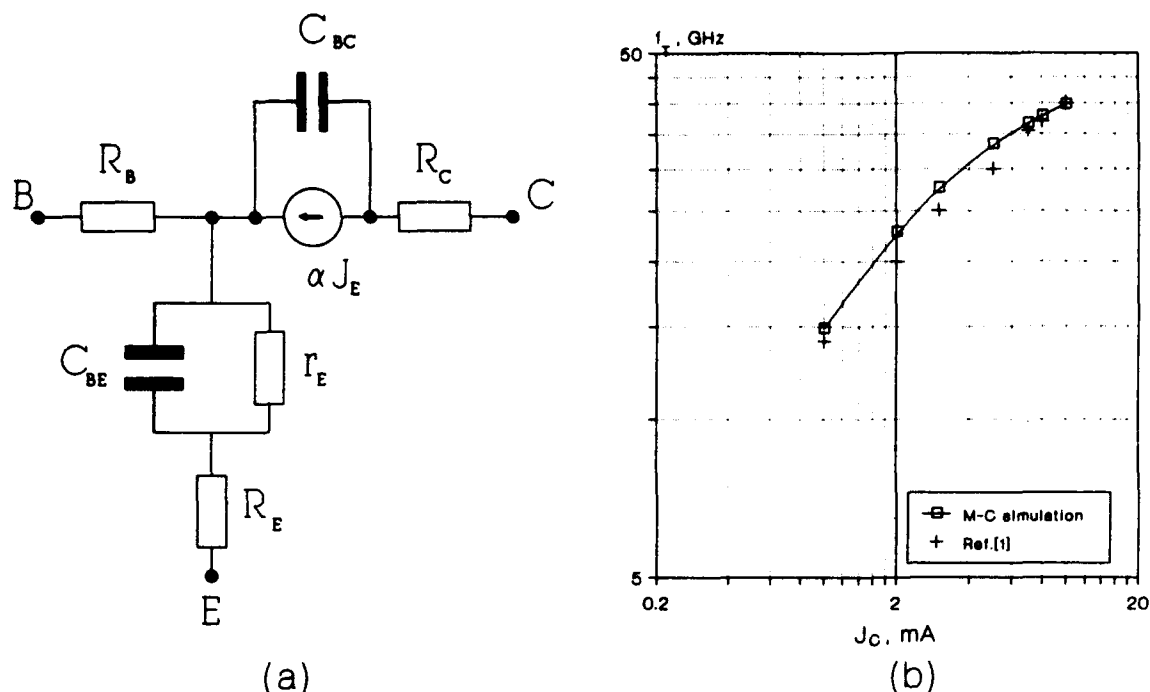


Fig.1 Small-signal equivalent circuit (a) and cut-off frequency f_T as a function of the collector current J_C (b).

$$\beta(\omega) \cdot \gamma_C(\omega) = \frac{F[J_C(t)]}{F[J_E(t)]}, \quad (3)$$

where F denotes the Fourier transform. Cut-off frequency f_T is defined from the analysis of the $h_{21}(\omega)$ computed by (3) as a frequency where $|h_{21}(\omega)| = 1$. It should be noted that proposed method is also appropriate to calculate other HBT high-frequency characteristics such as power gain, stability factor, scattering matrix and so on.

To verify proposed Monte Carlo particle model and method of evaluation of the HBT high-frequency performance the submicrometr HBT with the same structure as in [1] was simulated. The comparison of the simulated results with the experimentally observed results have shown high accuracy and validity of the model (see Fig.1).

III. Results and Discussions

A series of *AlInAs/GaInAs* HBTs with arrange of base thickness W_B from $0.02\mu\text{m}$ to $0.4\mu\text{m}$ and doped at $1.5 \cdot 10^{19}\text{cm}^{-3}$ was simulated at the temperature 300K to investigate the base thickness effect on the electron transport regime in the base.

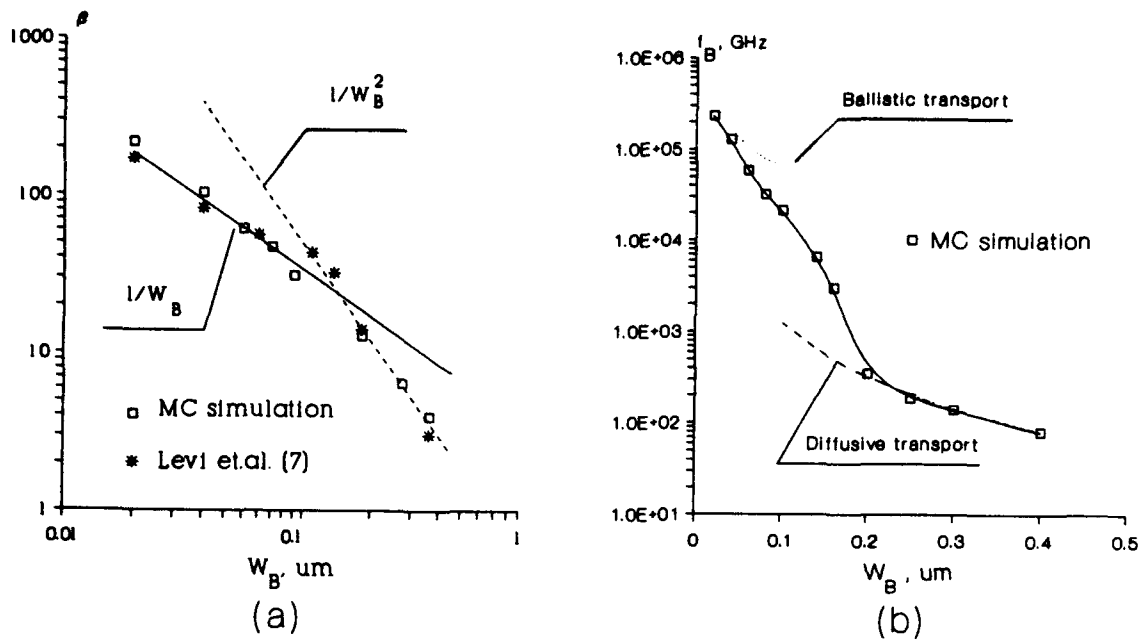


Fig.2 Common emitter current gain β (a) and base cut-off frequency f_B (b) as a function of the base thickness W_B .

The HBT common emitter current gain β dependence as a function of the base region thickness is shown in Fig.2. The analysis of this figure shows that $\beta(W_B)$ is proportional $1/W_B$ (corresponds to ballistic transport) for $W_B < 0.1 \mu\text{m}$ and is proportional $1/W_B^2$ (corresponds to diffusive transport) for $W_B > 0.1 \mu\text{m}$.

The base cut-off frequency $f_B = \frac{1}{2\pi\tau_B}$, where τ_B is a base delay time, dependence as a function of the base layer thickness is presented in Fig.2. Solid curve in the figure presents the dependence calculated by the numerical simulation (method based on the Fourier decomposition of the non-steady-state base current was used in present work to estimate the τ_B). Dashed curve presents the dependence calculated for diffusive transport and dotted curve corresponds to ballistic transport. It is seen that the ballistic transport take place only for $W_B \leq 0.06 \mu\text{m}$ and pure diffusive transport occurs while the base thickness is more than $0.2 \mu\text{m}$. For $0.06 \mu\text{m} \leq W_B \leq 0.2 \mu\text{m}$ the electrons transport in the base region is far from equilibrium but is neither pure ballistic nor pure diffusive. In this situation correct estimation of the base delay time is possible only by numerical simulation.

References

1. Y. Yamauchi and T. Ishibashi, "Electron velocity overshoot in the collector depletion layer of AlGaAs/GaAs HBT's," IEEE Electron Device Letters, Vol. EDL-7, no.12, pp. 655-657, 1986.
2. T. Ishibashi and Y. Yamauchi, "A possible near-ballistic collection in an AlGaAs/GaAs HBT with a modified collector structure," IEEE Trans. Electron Devices, Vol. ED-35, pp. 401-404, 1988.
3. N.A. Bannov, V.I. Ryzhii and A.A. Sviatchenko, "Computer modeling of the nonstationary electron processes in $n - p^+ - n$ -heterojunction bipolar transistors," Doklady Akademii Nauk SSSR, Vol. 287, pp. 1368-1373, 1986.
4. J. Hu, K. Tomizawa and D. Pavlidis, "Transient Monte Carlo analysis and application to heterojunction bipolar transistor switching," IEEE Trans. Electron Devices, Vol. ED-36, pp. 2138-2144, 1989.
5. G. Khrenov, V. Ryzhii and S. Kartashov, "Simulation of HBTs and HIGFETs based on AlGaAs/GaAs heterostructure," Tech. digest of 1991 Int. Conf. on VLSI and CAD, Seoul, Korea, pp. 215-218, 1991.
6. C.M. Maziar, M.E. Klausmeier-Brown and M.S. Lundstrom, "A proposed structure for collector transit-time reduction in AlGaAs/GaAs bipolar transistors," IEEE Electron Dev. Lett., Vol. EDL-7, pp. 483-485, 1986.
7. A.F.J. Levi, B. Jalali et al., "Vertical scaling in heterojunction bipolar transistors with nonequilibrium base transport," Appl. Phys. Lett, Vol. 60, pp. 460-462, 1992.

Scaled Ensemble Monte Carlo

A. M. Krizan¹ and R. P. Joshi²

¹Dept. of Elec. and Comp. Engineering,
State University of New York at Buffalo,
Buffalo, New York 14260 USA

²Dept. of Elec. and Comp. Engineering,
Old Dominion Univ.,
Norfolk, Virginia, 23529 USA

Abstract

We introduce a scaled ensemble Monte Carlo (SEMC) technique, useful for obtaining statistically significant results far into the low-density tails of carrier distributions. Standard approaches for studying this regime (weighted EMC) are based on determining single-particle distribution functions using a nonuniform trajectory sampling; the new technique instead stimulates a scaled distribution, with an energy-dependent scaling factor, but uses ordinary EMC weighting. This technique is flexible and simple to code. We display some results for bulk GaAs, with densities ranging over seven orders of magnitude, using only 10,000 particles.

I. General System Description

Our treatment is in the classical regime, in the absence of significant interparticle correlations, so the system is described formally by the usual time-dependent single-particle distribution function $f(\mathbf{r}, \mathbf{p}; t) = f(\mathbf{x}; t)$ [$\mathbf{x} = (\mathbf{r}, \mathbf{p})$ is a phase space coordinate]. The distribution function obeys the Liouville equation:

$$\frac{\partial f}{\partial t} = \{H, f\} + \left(\frac{\partial f}{\partial t}\right)_{\text{coll}}, \quad (1)$$

where $H = H(\mathbf{x})$ is the time-independent single-particle Hamiltonian, and $\{\cdot, \cdot\}$ is the Poisson bracket. We make the usual assumptions that collisions or scattering events take place on time scales much shorter than the time between collisions, so effects such as collision broadening can be neglected. Further, the scattering events are in fact approximated as instantaneous, so that intracollisional field effects can also be ignored. Under these assumptions, all important sources of potential and phonon scattering can be written in the form

$$\left(\frac{\partial f(\mathbf{x}; t)}{\partial t}\right)_{\text{coll}} = -\Gamma(\mathbf{x}; t) f(\mathbf{x}; t) + \int \Gamma(\mathbf{x}, \mathbf{x}'; t) f(\mathbf{x}'; t) d\mathbf{x}' . \quad (2)$$

II. Scaling Formalism

The approach developed depends primarily on the observation that an energy- and time-dependent scaling factor commutes with noncollisional term on the right-hand side:

$$s(H, t) \{H, f\} = \{H, s(H, t) f\} . \quad (3)$$

Thus, we define a *scaled* distribution function defined by

$$\hat{f}(\mathbf{x}; t) \equiv s(H(\mathbf{x}), t) f(\mathbf{x}; t) . \quad (4)$$

The principal advantage of multiplying by an energy-dependent factor arises from counting statistics. If we are interested in the distribution function in some region about the phase space point \mathbf{x} , we consider as a function of time the number of simulation particles N in the vicinity of that point. The fractional error in \hat{f} is then $1/\sqrt{N}$, which is the fractional error in $f \equiv s^{-1} \times \hat{f}$ as well. For regions of low phase-space density, N is proportional to $f(\mathbf{x};t)$. Thus, for example, to examine regions where the phase space density is down by seven orders of magnitude from the maximum, one needs many times of ten million simulation particles to keep the error from exceeding the estimate.

A standard solution to this problem — weighted EMC — is to define a region of interest in phase space, and to perform multiple simulations of those few particle trajectories which enter it. Usually, this is coded in a way that sharply distinguishes high- and low-density regions. If the density falls smoothly, there is no efficient place to draw the boundary between these two regions. Other, more subtle approaches have also been used [1].

The present approach is based upon scaling the distribution function that is simulated by EMC, rather than upon weighting the EMC simulation of an unscaled distribution function. The choice of s determines the trajectory density. In principal, these two approaches may be equivalent in particular cases. However, a weighted EMC eliminates the usual identification between individual initial particles and individual trajectories sampled (in what is the Monte Carlo integration of the Boltzmann equation). As a result, one loses the intuitive simplicity of regarding sampled trajectories as individual particles of a large ensemble. In a scaled EMC, on the other hand, one preserves a one-to-one correspondence of initial condition to trajectory, and it remains possible to regard the trajectories sampled as the actual paths of individual particles. As we describe below, however, in order to redistribute the statistical sampling weight, one pays the price that the trajectories do not follow the paths of ordinary particles.

III. Time-dependent scaling

Particle-number conservation imposes an important constraint on how $s(H(\mathbf{x}),t)$ is allowed to be chosen. By appropriate normalization, the total number N of simulation particles in the simulation of \hat{f} is the integral of the scaled distribution function:

$$\hat{N}(t) = \int \hat{f}(\mathbf{x}) d\mathbf{x} = \int s(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} . \quad (5)$$

If s is chosen to emphasize high-energy regions which have low density, then during a relaxation, thermalization will cause a transfer of (real) electrons to lower-energy regions where s is smaller. If s is not allowed a time-dependence to compensate, the total number of simulation particles of \hat{f} must decrease — degrading the statistics in \hat{f} . Conversely, a heating mechanism would increase the number of simulation particles, improving statistical precision but possibly requiring undesirable computational expense. By allowing s to have a time-dependence, we accomplish in the time domain what the energy dependence of s accomplishes for the energy domain: reduce variations in particle number so that fractional errors can be kept at an acceptable level throughout the region of interest, with the least computational effort.

We chose a simple form for the joint energy-and-time dependence of s : we let them be independent factors. This can be written

$$s(H,t) = \exp(\alpha(t) + \gamma(H)) . \quad (6)$$

Furthermore, we let the function $\gamma(H) = -H / k_B T^{\text{eff}}$. This is appropriate for distributions f which are approximately characterized by effective temperatures below T^{eff} [Higher temperatures lead to normalization problems with \hat{f} .] This choice is also convenient

computationally: a linear function γ implies that inelastic scattering rates as well as elastic rates, are space-position-independent, and can be stored efficiently in look-up tables.

\hat{f} then obeys a modified Liouville equation

$$\frac{\partial \hat{f}}{\partial t} = \{H, \hat{f}\} + \left(\frac{\partial \hat{f}}{\partial t}\right)_{\text{coll}}, \quad (7)$$

in which the "collision" or scattering term is defined by

$$\left(\frac{\partial \hat{f}}{\partial t}\right)_{\text{coll}} - \frac{1}{s} \left(\frac{\partial s}{\partial t}\right) \hat{f} = s \left(\frac{\partial f}{\partial t}\right)_{\text{coll}}. \quad (8)$$

Equations (2) and (8) specify completely the modified scattering term $(\partial \hat{f} / \partial t)_{\text{coll}}$. However, in order to implement a Monte Carlo time-evolution, one must determine scattering rates for the scaled problem which are analogous to the out-scattering rates $\Gamma(\mathbf{x}; t)$ and the in-scattering rates $\Gamma(\mathbf{x}, \mathbf{x}'; t)$ of the unscaled problem. There is some freedom in way this is done. One well-known degree of freedom is associated with self-scattering:

$$\left\{ \begin{array}{l} \Gamma(\mathbf{x}; t) \rightarrow \Gamma(\mathbf{x}; t) + \Delta\Gamma(\mathbf{x}; t) \\ \Gamma(\mathbf{x}, \mathbf{x}'; t) \rightarrow \Gamma(\mathbf{x}, \mathbf{x}'; t) + \Delta\Gamma(\mathbf{x}; t) \delta(\mathbf{x} - \mathbf{x}') \end{array} \right\}, \quad (9)$$

where ordinarily $\Delta\Gamma$ is chosen to make the total out-scattering a positive constant. This is a kind of gauge transformation, in which the physically-significant total scattering rate is fixed, while unobservable components of the in- and out-scattering rates make the numerical implementation tractable. A more general kind of gauge transformation is made in the scaled EMC approach, leading to off-diagonal in-scattering rates of the form

$$\hat{\Gamma}(\mathbf{x}, \mathbf{x}') = G(\mathbf{x}, \mathbf{x}') + G(\mathbf{x}) \frac{1}{N} \hat{f}(\mathbf{x}) \quad \text{for } \mathbf{x} \neq \mathbf{x}'. \quad (10)$$

The second term on the right-hand side leads to a kind of attractive interparticle scattering. This performs a rôle similar to that of trajectory iteration in weighted-EMC approaches: simulation particles entering critical regions are given greater weight, and are effectively caused to perform multiple traversals. However, in SEMC this weighting is implemented smoothly, rather than abruptly at the boundary of a region of interest, and it is accomplished with a fixed number of particles undergoing essentially ordinary scattering.

IV. Simulations

We have applied the SEMC technique to bulk GaAs semiconductor at 300 K. We used parameters (deformation potentials, phonon energies, band structure, etc.) that have been confirmed empirically in previous simulations [2]. We have specifically neglected hot phonon effects and Coulomb scattering, so we have essentially modeled the electrons in intrinsic GaAs with weak laser excitation. The system was allowed to relax for 10 ps from the initial laser excitation, and the result plotted below were obtained as an average of the distribution function during the last 0.5 ps (i.e., an average was performed of results of the last 100 5-ps observation times). Because of the averaging procedure used, there is a kind of local averaging of the density, so that the standard deviation of the results from a smooth fit underestimates the uncertainty of the simulation results.

Figure 1 shows the density as a function of energy of electrons in the central (Γ) valley. An ordinary (unscaled, unweighted) EMC simulation is compared with the scaled simulation using $T_{\text{eff}} = 400$ K. *Both of the simulations use the same (constant) number of simulation particles: 10^4 .* In this valley, the SEMC simulation appears to be *less* smooth (i.e., to have greater

statistical uncertainty) at low energies. This is in fact correct: the price we pay for increasing the number of simulation points at higher energies is to reduce the number at low energies, with correspondingly poorer statistics. At higher energies, when the densities become smaller than 10^{-4} , the usual EMC becomes completely unreliable, since results correspond to single simulation particles in a bin. The results of the two EMC simulations are consistent (in the sense that their difference is not statistically significant), as they must be, since they are mathematically equivalent in the large- N limit. The equivalence at high energies is clearly due to the large error bars in the unscaled EMC (not shown) which completely cover the range of densities to zero. There is the *appearance* that the unscaled EMC predicts a systematically higher density at high energy, since all of the high-energy points plotted are above the corresponding SEMC points plotted. There are two reasons for this appearance. One is the local (time) averaging discussed, which has the effect of smoothing a random statistical fluctuation into an apparent systematic one. The other is that on the semilogarithmic scale, the points where EMC predicts zero density (and lie *below* the SEMC simulation) cannot be plotted.

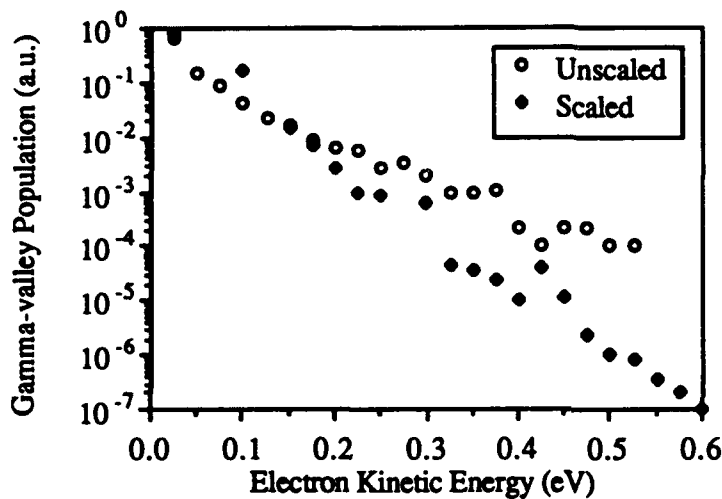


Figure 1. Central conduction valley population density.

Figures 2 and 3 show the same comparison as figure 1, for the L- and X-valleys, respectively. The thresholds at zero kinetic energies correspond to satellite band minima many times $k_B T$ above the Γ -valley minimum, and so the regime of poor statistics is reached more quickly. The SEMC in each case shows the smooth exponential fall-off of a nearly equilibrated system.

V. Conclusion

We have shown that a scaled EMC simulation can be implemented to significantly improve the statistics over broad ranges of low real densities.

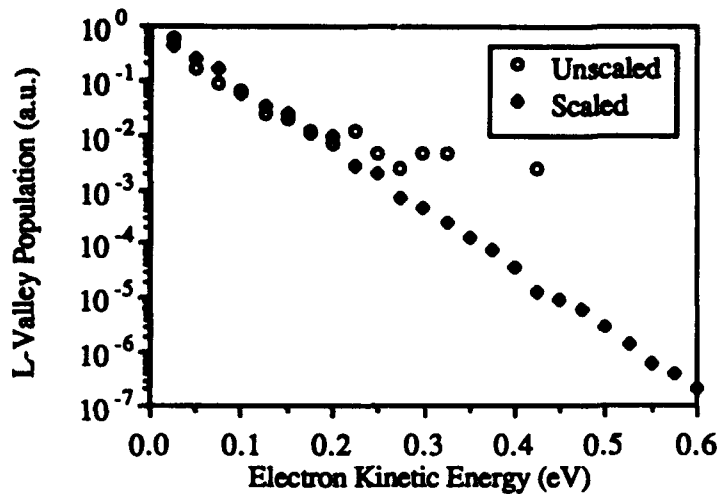


Figure 2. L-valley population density.

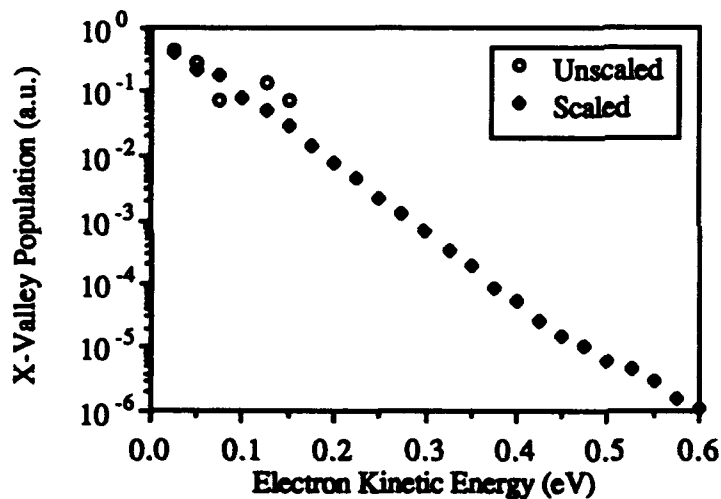


Figure 3. X-valley population density.

References

- [1] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation*, Springer-Verlag, 1989.
- [2] M.-J. Kann, A. M. Kriman and D. K. Ferry, "Effect of Electron-Electron Scattering on Intervalley Transition Rates of Photoexcited Carriers in GaAs," *Phys. Rev. B* Vol. 41, No. 18, pp. 12659-12665, 15 June 1990.

ACOUSTIC PHONON CONTROLLED TRANSPORT IN LOW DIMENSIONAL STRUCTURES

V. Mitin, R. Mickevičius, and N. Bannov
Department of Electrical and Computer Engineering
Wayne State University, Detroit, MI 48202

Abstract

The rates of electron scattering by acoustic phonons in low dimensional structures are calculated. It is shown that due to uncertainty of momentum conservation in low dimensional systems the acoustic phonon scattering becomes essentially inelastic in contrast to that in bulk materials. We propose simple procedure for incorporation of inelastic acoustic phonon scattering into a Monte Carlo technique. It is demonstrated that electron low-field mobility at temperatures less than 100 K calculated within elastic approximation is greatly underestimated. We have also demonstrated that velocity-field dependence in QWIs has a superlinear region. This superlinear region is associated with decline in acoustic phonon scattering efficiency as the electrons get heated.

I. Introduction

Low dimensional (LD) semiconductor structures are now widely recognized as a very promising basis for future technological applications. The electron transport in LD structures at low temperatures is controlled by acoustic phonon scattering. However, so far there is a considerable gap in understanding of some essential aspects of LD electron scattering by acoustic phonons. The elasticity of electron-acoustic phonon scattering is a commonly used approximation [1-4]. A closer look at this mechanism shows that electron scattering by acoustic phonons in quantum structures becomes essentially inelastic and becomes effective mechanism of energy dissipation [5]. This is due to the fact that the momentum conservation in LD structures for electron-acoustic phonon systems is not preserved.

We present results of calculations of electron scattering by acoustic phonons in 1D quantum wires (QWIs) and 2D quantum wells (QWs), where we consider inelasticity of this scattering in full detail. We develop procedure for inclusion of this inelasticity into a Monte Carlo technique. Monte Carlo simulation results are obtained for GaAs QWIs.

II. Scattering Rates

The rate of electron scattering from the state \mathbf{k} to the state \mathbf{k}' with the assistance of deformation acoustic phonon with wave-vector \mathbf{q} is given by

$$W(\mathbf{k}, \mathbf{k}', \mathbf{q}) = \frac{\pi E_a^2 q^2}{\rho V \omega_q} \left(N_q + \frac{1}{2} \pm \frac{1}{2} \right) \delta(\epsilon_k - \epsilon_{k'} - \Delta\epsilon \mp \hbar\omega_q), \quad (1)$$

where V is the principal volume of the crystal, E_a is the deformation acoustic potential, ρ is the density of the material, $\omega_q = uq$ is the long wave approximation of acoustic phonon dispersion, where u is a sound velocity in the material, and $\Delta\epsilon$ is the energy separation

between initial and final subbands. Here and everywhere below upper sign corresponds to emission and lower sign to absorption of acoustic phonon. We will consider a rectangular quantum structure with infinitely deep potential well embedded into another material with similar elastic properties so that acoustic phonons can penetrate through the interface between these materials. The major difference between bulk and LD electron scattering comes from the overlap integral $I = \langle \psi_{k'} | e^{iqz} | \psi_k \rangle$. Integrating over the volume of the quantum structure and neglecting umklapp processes we get

$$|I_{2D}|^2 = G_{n,n',q_z} \delta_{k_{||} \mp q_{||}, k'_{||}},$$

$$|I_{1D}|^2 = G_{j,j',q_z} G_{l,l',q_z} \delta_{k_z \mp q_z, k'_z}, \quad (2)$$

$$G_{l,l',q_z} = \frac{2[(2\pi)^2 q_z L_z l l']^2 [1 - (-1)^{l+l'} \cos(q_z L_z)]}{[(q_z L_z)^4 - 2\pi^2 (q_z L_z)^2 (l^2 + l'^2) + \pi^4 (l^2 - l'^2)^2]^2}. \quad (3)$$

where n is subband index for a QW; j and l are subband indices for a QWI due to size quantization in y and z directions, respectively. Primed indices denote the subband of the final state. The Kroenecker delta-function reflects momentum conservation in unquantized direction(s). The form-factor G is responsible for the uncertainty of momentum conservation in perpendicular to the quantum structure direction(s).

The total scattering rate from the state k of the initial subband to any state of the final subband,

$$\lambda(k) = \frac{V}{(2\pi)^3} \sum_{k'} \int dq W(k, k', q). \quad (4)$$

The complexity of the form-factor does not allow one to get analytical results without certain simplifications. Let us consider now separately QWs and QWIs. For numerical estimates we use parameters of GaAs quantum structure embedded in AlAs at $T=30$ K.

(a) Quantum Wells

The scattering rate of 2D electrons by acoustic phonons obtained from (7) for electrons on subband n with wave vector $k_{||}$ scattered to subband n' can be rewritten in the following form:

$$\lambda_{n,n'}(k_{||}) = \frac{E_a^2 m k_B T}{2 L_z \rho \hbar^3 u^2} \mathcal{J}_{n,n'}(k_{||}), \quad (5)$$

where $\mathcal{J}_{n,n'}(k_{||})$ is a dimensionless function of the order of unity, given by

$$\begin{aligned} \mathcal{J}_{n,n'}(k_{||}) = & \int_0^{+\infty} \frac{L_z dq_z}{\pi} \int_0^\pi \frac{d\varphi}{\pi} \mathcal{N} \left(\frac{\hbar u q}{k_B T} \right) G_{n',n,q_z} \times \\ & \times \left| 1 \pm \frac{m u}{\hbar k'_{||}} \frac{k'_{||} - k_{||} \cos(\varphi)}{\sqrt{k_{||}^2 + k'_{||}^2 - 2 k_{||} k'_{||} \cos(\varphi) + q_z^2}} \right|^{-1} \theta(k_{||}, \varphi, q_z, n, n'). \end{aligned} \quad (6)$$

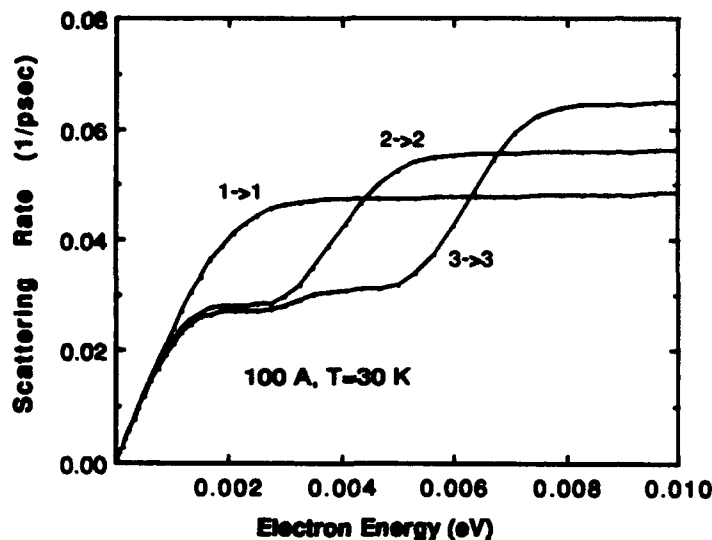
We use the following notation in formula (6): φ is an angle between electron wave vector before scattering $k_{||}$ and electron wave vector after scattering $k'_{||}$. An absolute value of the in-plane electron wave vector after scattering, $k'_{||}$, is a solution of the energy conservation equation. The function $\mathcal{N}(x) = x/(e^x - 1) + x/2 \pm x/2$ in eq. (6) is a dimensionless

function associated with phonon occupation number, $\theta(k_{||}, \varphi, q_z, n, n')$ is an integer-value function which is equal to a number of the roots of the energy conservation equation.

For the elastic approximation $\hbar u q \rightarrow 0$ we obtain the following expression:

$$\mathcal{J}_{n,n'}(k_{||}) = (1 + \frac{1}{2}\delta_{n',n})\Theta(\epsilon_n - \epsilon_{n'} + \frac{\hbar^2 k_{||}^2}{2m}),$$

where $\Theta(x)$ is a unit step-function. The elastic approximation fails in the region of low electron energy which amount for 1–5 meV for GaAs QWs. Mathematical complexity of the problem does not allow one to solve it analytically. Figure 1 demonstrates the energy dependence of the acoustic phonon emission rate in three lowest subbands (intrasubband transitions) of the 100 Å thick QW for $T=30$ K, obtained by direct numerical solution of eqs. (5) and (6). Instead of single-step-wise function which follows from the elastic approximation, the emission rate is a smooth function of electron energy. The complex structure of the form-factor for transitions within upper subbands results in several plateau regions in the energy dependence of the emission rate.



(b) Quantum Wires

Let us assume that $q_z \ll q_T$, where $q_T = \sqrt{q_y^2 + q_x^2}$ is a transverse component of phonon wavevector and that $\hbar u q \ll k_B T$. To meet both above assumptions it is required that $\hbar u q_{Tmax} \ll k_B T$, where $q_{Tmax} = \pi \sqrt{(j+j')^2/L_y^2 + (l+l')^2/L_x^2}$ is roughly the maximum transverse component of phonon wavevector where the form-factor is non-zero. Then after summation of eq. (7) over k' and integration over q_z we get

$$\lambda(\epsilon) = \frac{E_a^2 k_B T}{4\pi^2 \rho u^2 \hbar^2} \sqrt{\frac{m^*}{2}} \int \frac{dq_y dq_x G}{\sqrt{\epsilon \mp \hbar u q_T - \Delta\epsilon}} \quad (7)$$

Here G stands for $G_{j,j',\epsilon_y} G_{l,l',\epsilon_x}$. The integration limits for absorption rate are imposed by the form-factor which tends to zero above q_{Tmax} . The uncertainty of momentum conservation is roughly defined by $\hbar q_{Tmax}$, i.e., it increases with decreasing the transverse dimensions of the QWI. As the electron energy tends to zero the absorption rate saturates since q_{Tmax} is not equal to zero.

Let us consider the asymptote of very large electron energies $\epsilon - \Delta\epsilon \gg \hbar u q_{Tmax}$. In this case we may neglect phonon energy (elastic approximation). Note that for this asymptote emission and absorption rates are equal. Then (7) turns into:

$$\lambda(\epsilon) = \frac{E_a^2 k_B T}{\rho u^2 \hbar^2 L_y L_x} \sqrt{\frac{m^*}{2(\epsilon - \Delta\epsilon)}} (1 + \frac{1}{2}\delta_{j,j'}) (1 + \frac{1}{2}\delta_{l,l'}) \quad (8)$$

Thus the large energy asymptote diverges as $\epsilon^{-1/2}$ function of electron kinetic energy after scattering, the same as 1D density of states function. The large energy asymptote is extended to low energy region in [1-4] to evaluate electron mobility. The scattering rate within this elastic approximation is inversely proportional to a QWI cross section $L_y \times L_z$.

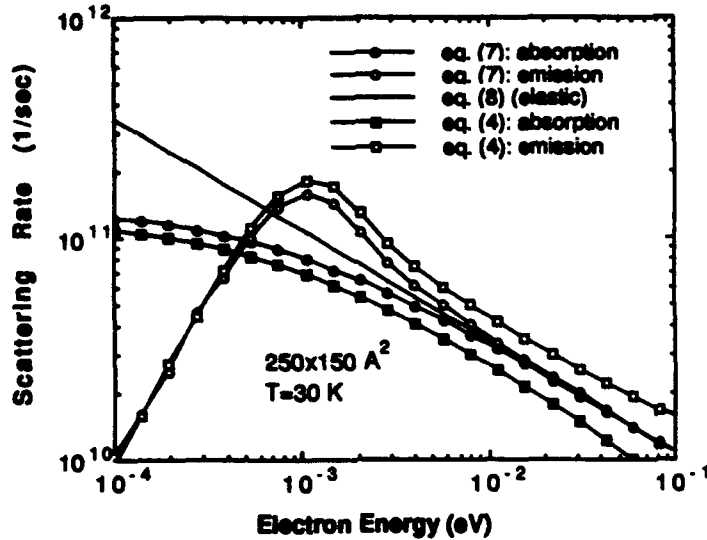


Figure 2 demonstrates the differences between the scattering rate dependencies on electron energy calculated for QWI with cross section $250 \times 150 \text{ \AA}^2$ within three different approaches: elastic approximation of eq. (8), the approach of eq. (7), and the direct numerical integration of eq. (4) without any simplifications. The correct inelastic treatment of acoustic phonon scattering yields qualitatively different scattering rates. Emission and absorption rates within the inelastic approach are considerably different at low energies.

There is no divergency of the scattering rates within inelastic approach. The energy $\hbar\omega_{qT_{max}}$ given by uncertainty of momentum conservation defines the position of the maximum on emission curve and coincides with the characteristic energy of acoustic phonon interacting with electrons.

III. Results of the Monte Carlo Simulation

The necessity to allow for inelasticity of acoustic phonon scattering in Monte Carlo simulations complicates the procedure of choice of the final states for scattered electron. We have developed very efficient procedure for random selection of acoustic phonon energy involved in the scattering. The essence of this procedure is that we first numerically perform von Neumann procedure for a set of random numbers and tabulate the phonon energy as a function of a random number. For a QWI we have solved the following equation:

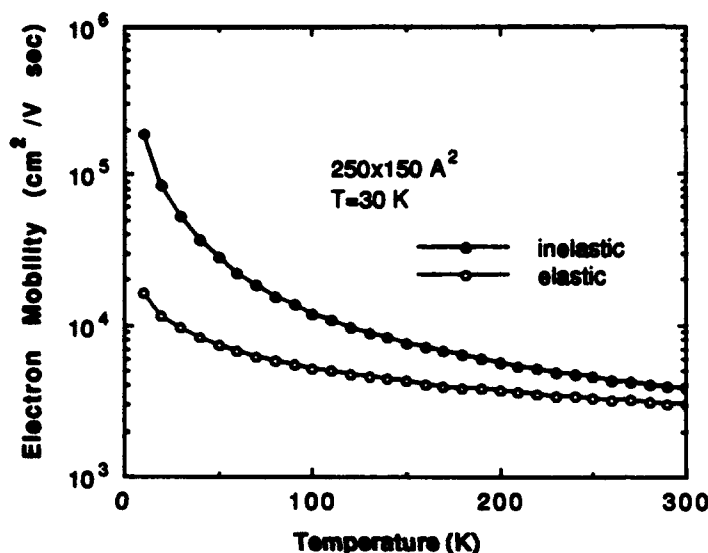
$$\tau \int_0^{q_{T_{max}}} dq_T F(q_T) = \int_0^{q_T^*} dq_T F(q_T), \quad (9)$$

with respect to q_T^* for a set of 100 random values of τ ranging from 0 to 1. Here the function $F(q_T)$ is an integrand of eq. (7) and represents the scattering probability dependence on the transverse components of the phonon wavevector. Since q_T is directly related to the phonon energy within approximation $q_s \ll q_T^1$, one can find the desired phonon energy for each value of τ . A table of such values has been calculated for a set of electron energies

¹Within more accurate approach where q_s is not neglected, the acoustic phonon energy is uniquely related to q_s , so that this dependence can be used to generate the acoustic phonon energy [5].

for emission and absorption of acoustic phonon separately. Then the choice of phonon energy involved into a scattering event is just the generation of random number r and the selection of the corresponding phonon energy value from the appropriate table. This procedure essentially speeds up the Monte Carlo simulation.

To test the scattering rates and the Monte Carlo procedure we have simulated the electron transport at low lattice temperatures. Monte Carlo simulations yield very accurate thermal equilibrium energy and distribution function for electron system indicating that acoustic phonon scattering is treated correctly. The elastic scattering model fails to yield good equilibrium energy and distribution function. Figure 3 demonstrates low field



electron mobility in the QWI with cross section $40 \times 40 \text{ Å}^2$ as a function of lattice temperature. One can see that electron mobility calculated within elastic approach is significantly underestimated in the low temperature region. The reason for this is obvious: an overestimation of acoustic phonon scattering rate (especially emission rate) at low energy region. It must be noted that in thicker QWIs the discrepancy between low field electron mobilities obtained within elastic and inelastic models is less pronounced because inelasticity

of acoustic phonon scattering is weaker in thick QWIs.

We have also observed superlinear region on velocity-field dependence associated with reduction of acoustic phonon scattering efficiency as the electron system gets heated. This superlinear region appears only for rather thick QWI. In thin QWIs (cross section of the order of $40 \times 40 \text{ Å}^2$) the acoustic phonon scattering is so strong and inelastic that it prevents electron heating up to very high electric fields. The superlinearity on velocity-field dependence is greatly overestimated within elastic approximation since elastic scattering does not prevent electron runaway from the low energy region. As a result, the velocity-field dependences calculated within elastic and inelastic approaches are considerably different.

Acknowledgement: This work is supported by the NSF and by ARO.

References

- [1] F. Comas, C. Trallero Giner, and J. Tutor, *phys. stat. solidi (b)* 139, 433 (1987).
- [2] V. K. Arora and M. Prasad, *phys. stat. solidi (b)* 117, 127 (1983).
- [3] V. K. Arora and A. Naeem, *Phys. Rev. B* 31, 3887 (1985).
- [4] M. P. Chamberlain, D. Hoare, R. W. Kelsall, and R. A. Abram, *Semicond. Sci. Technol.* 7, B45 (1992).
- [5] R. Mickevičius and V. Mitin, submitted to *Phys. Rev. B*.

Highly Efficient Full Band Monte Carlo Simulations

R. Kent Smith and Jeff Bude

AT&T Bell Labs, Murray Hill NJ, 07974-0636

Abstract

We present a full band monte carlo algorithm based on phase-space simplexes which has all of the advantages of analytical band monte carlo while preserving the accuracy of a full band structure. An adaptive, contour-aligned grid algorithm is used to represent the energy band structure within the irreducible wedge and is calculated once for each semiconductor band. The complexity of generating the adaptive grids in phase-space is more than compensated for by the simplicity and efficiency introduced into the equations of motion and final state selection, which are treated exactly within the physics of the model. Consequently, our method confines simulation error to statistical error and a consistently bounded discretization error determined only by the choice of phase-space grid and can be set to a user-defined tolerance. Results using this method show at least an order of magnitude improvement in performance over previous full band codes.

Monte Carlo simulation has earned an important place in semiconductor transport simulation because it offers a practical way of solving the full Boltzmann equation - essential for a rigorous understanding of non-stationary/high-field transport which includes impact ionization, hot carrier dynamics, and velocity overshoot. Naturally, the more rigorously based the physics in the Monte Carlo simulation, the more accurate and detailed the solution. The numerical solution of the Boltzmann equation is obtained by integrating particles along trajectories in phase space. The momentum of each particle is abruptly changed when scattering events occur.

One of the most important elements in establishing an accurate physical model is the inclusion of a full band structure (FB) which, for example, can be calculated from empirical pseudo-potential or local density functional theory. Unfortunately the inclusion of a full band structure considerably complicates the MC simulation. First, the particle velocity is not simply related to \vec{k} but is determined by the gradient of the energy band. Failure to integrate the equations of motion exactly will result in unphysical gains or losses in the particle energy. Secondly, the scattering rates conserve energy which, for simple dispersionless phonon models, restricts the final

state momentum, \vec{k}' , to lie on energy isosurfaces. For a general band structure, these surfaces may be extremely complicated requiring elaborate searches of the Brillouin Zone for the final state momentum, \vec{k}' .

The computation complexity of MC simulations may be significantly reduced if an analytic expression is used for the band structure. Isotropic polynomial approximations to the energy bands enable efficient and accurate calculations of the free flight trajectories and final states momentum selection. Analytical bands (AB) are rigorously justified for low field transport since the carriers remain close to the band minimum, where the bands are parabolic. Although AB models are computationally efficient, accuracy is sacrificed, at high energies, where simple band approximations are difficult to construct.

We propose a FB MC algorithm [2, 3], based on simplexes, which has the computational advantages of the faster AB simulators, while preserving the physics contained in the full band structure. In k -space, the irreducible wedge is represented by a contour-aligned tetrahedral grid. An adaptive algorithm is developed to approximate each energy band by a piecewise linear polynomial to a pre-defined accuracy. Within each element, the equations of motion may therefore be integrated exactly, thus conserving energy at each time step. Furthermore, the energy isosurfaces for the entire Brillouin Zone, may be readily calculated providing an accurate final k -state selection. As a result, the simulation error to the statistical error inherent in MC, and a consistently bounded discretization error deriving from the choice of grid in phase-space.

Our algorithm represents phase space as a cartesian product of two simplex grids, $T_k \times T_x$. In k -space, only the irreducible wedge, defined by the 48 elements of the point group pertaining to cubic semiconductors is triangulated. A rotation matrix, Q , is assigned to each particle rotates the position in the irreducible wedge to its proper place in the full Brillouin zone. In both spaces, piecewise linear polynomial approximations are used to represent the electrostatic potential and each energy band.

Since the equation of motions are linear, particle trajectories may be computed exactly. Pointers to adjacent elements are used to facilitate motion through the grid. Neighbor pointers are also constructed at the phase space boundaries, where particles may be either transmitted or reflected. Element pointers at reflected surfaces point to themselves, where as transmitted particles reenter the phase space grid at some other point. In either case, the simulation particles remain confined to the solution domain and enough information is retained to reproduce the entire phase space.

The selection of the final state momentum is also exactly computed for piecewise linear polynomial approximations. Since ∇E is a constant, the constant energy surfaces are defined by the intersection of a plane with the tetrahedron. An efficient, single parameter, quadratic interpolation procedure may be developed when the vertices of the tetrahedron are constrained to lie between two constant energy surfaces. Each tetrahedra classified by the number of vertices lying on an given energy contour, E_i . The density of states is then characterized by the single interpolation parameter,

$\phi(E)$, and given by

$$\mathcal{D}^{(I)}(E) = \mathcal{D}_T^{(I)}(1 - \phi)^2 \quad (1)$$

$$\mathcal{D}^{(II)}(E) = \mathcal{D}_T^{(II)}\phi(1 - \phi) \quad (2)$$

$$\mathcal{D}^{(III)}(E) = \mathcal{D}_T^{(III)}\phi^2 \quad (3)$$

where $\phi(E) = (E - E_i)/(E_{i+1} - E_i)$ and $E_i \leq E \leq E_{i+1}$. The major advantage of this decomposition is that global quantities, obtained by summing over tetrahedra, may be interpolated in the same manner. The selection of the final state momentum is readily computed by proceeding down a k-space tree. The rotation matrix, Q , is first determined, then the tetrahedra type, and finally a single tetrahedron containing the final state. In addition to ϕ , two uniformly distributed random variables are sufficient to determine the barycentric coordinates within the tetrahedron. The final state momentum is then computed by a linear interpolation of the tetrahedron vertices, \vec{k}_i .

Since the trajectory calculations and the selection of the final state momentum are performed exactly, the only numerical error encountered in the solution of the discretized BTE is the statistical error associated with sampling the phase space. This error is bounded during each monte carlo run by monitoring the fluctuations in computed quantities. However, a second source of numerical error arises from approximating the continuous BTE with a discrete set of equations. Complete control of the numerical error is obtained by constructing meshes that provide a consistent bound on the discretization error.

We use an adaptive grid algorithm to construct meshes that represent functions in both k-space and x-space to a desired accuracy. These algorithms are based upon estimating the discretization error, ϵ , in approximating continuous functions by piecewise linear polynomials. In x-space, simple, inexpensive error estimates are constructed directly from Poisson's equation, [1]. These estimates provide both lower and upper bounds to the true discretization error, $0 \leq C_1|||\epsilon||| \leq |||\psi - \psi_h||| \leq C_2|||\epsilon|||$. For k-space grids, the discretization error is computed as deviations of the true band energies from piecewise linear approximations. The meshes are continually refined or unrefined until $||\epsilon|| \leq \delta$ where δ is a user specified error tolerance. By controlling the discretization error of these phase space grids, we can insure that numerical error associated with our MC simulations is bounded.

Several considerations are necessary to produce an acceptable, contour-aligned, tetrahedral grid for our monte carlo simulations. Unlike conventional adaptive refinement algorithms, each grid edge must lie either on an contour or connect two adjacent contours. Mesh refinement is accomplished by adding new energies contours rather than simply refining element edges. Sharp cusps are formed in the higher energy bands, due to band crossings, which produce discontinuous changes in $\nabla E(\vec{k})$. These features significantly effect the accuracy of any polynomial approximation to the energy bands. As with any mesh generation code, the grid should consist of reasonably shaped tetrahedra.

Table 1: Tetrahedra Grid Summary

Band	Vertices	Tetrahedra	Contours	Quality
1	349	1198	24	0.567
2	530	2268	19	0.611
3	414	1617	18	0.590
4	1343	6143	16	0.529
5	1658	8077	17	0.564
6	613	2431	13	0.456
7	419	1542	11	0.421

Critical points and ridge lines play an essential role the production of accurate contour-aligned grids. The role of critical points, \vec{k} such that $\nabla E(\vec{k}) = 0$, is twofold. Local extrema make excellent vertices, especially for coarse piecewise linear approximations. Also, saddle points describe how components of contours split and connect [4, 9, 8, 10]. The role of ridge lines, extrema in the curvature of the isosurface, is used to decompose the space between two contour surfaces. When a ridge line connects, say, a minimum to an adjacent saddle point, the "watershed" regions between ridge lines are empirically observed to provide a good starting decomposition.

The initial grid is generated from a skeleton of each energy band. The skeleton consists of a set of ridge lines and critical point energies that divide the wedge into several regions. Within each region, vertices are generated along contours that adhere to both the wedge boundaries and ridge lines. Curved region boundaries are represented by piecewise linear line segments with sufficient resolution to capture the geometry of the boundary. Contour-aligned grids are produced by refining grid edges. Edges that cross more then one contour are divided by adding a point at the midpoint contour energy. Edges that lie on a contour are refined either to reduce the discretization error or to improve the quality of the resultant tetrahedra. After all edges are divided at most once, the grid is then retriangled producing a new set of tetrahedra. The overall accuracy of the contour-aligned grids is controlled by selectively adding new energy contours when necessary.

The lowest seven silicon energy bands were computed using the nonlocal pseudopotential of Chelikowsky and Cohen [5]. The first 3 bands correspond to holes while the last 4 bands correspond to electrons. To easily accommodate interband transitions, a single set of energy contours was calculated for all bands. A error tolerance of $\delta_k = 0.02E(\vec{k})$ was used in generating all grids. The low energy spectrum was further selectively refined to produce an error if $\delta = 0.002\text{ev}$ near the minimum energy. As shown in Table 1, accurate, high quality, contour-aligned tetrahedral grids can be obtained with relatively few grid points. The grid for the lowest electron band in the x-y plane of the irreducible wedge is shown in figure 1.

The performance of our algorithm is compared with several existing monte carlo

Table 2: Performance Comparison

Code	Band Structure	CPU Time (YMP)	
		Time Step	Total
IBM	Full	5.0*	100.0*
Illinois	Full	1.5	25.0
BEBOP	Analytic	1.0	2.0
MMC	Full	1.0	1.0

* estimated

codes in Table 2. The test example consisted of simulating 10000 electrons in a constant electric field of 10^5 V/cm for a simulation time of 1 picosecond. To obtain a fair comparison, the full band Illinois code [11], the analytic band BEBOP code [12], and our full band code, MMC, were run on the same machine, a Cray YMP. The CPU times for IBM monte carlo code [7] were taken from published data [6]. These timings are qualitative since a different problem was solved on a different machine. Large time steps may be taken for our simulations and BEBOP, since the equations of motions and the final state energy selection are computed to the machine precision. Small time steps for the other full band codes are necessary to integrate the equations of motion to a reasonable numerical accuracy due to the general band structure. The significant increase in CPU time of the IBM code as compared to the Illinois code is largely attributed to the exhaustive time spent in the final state momentum selection. By any measure, our code combines the rigorous physics of full band simulations with the computational efficiency of analytical band models.

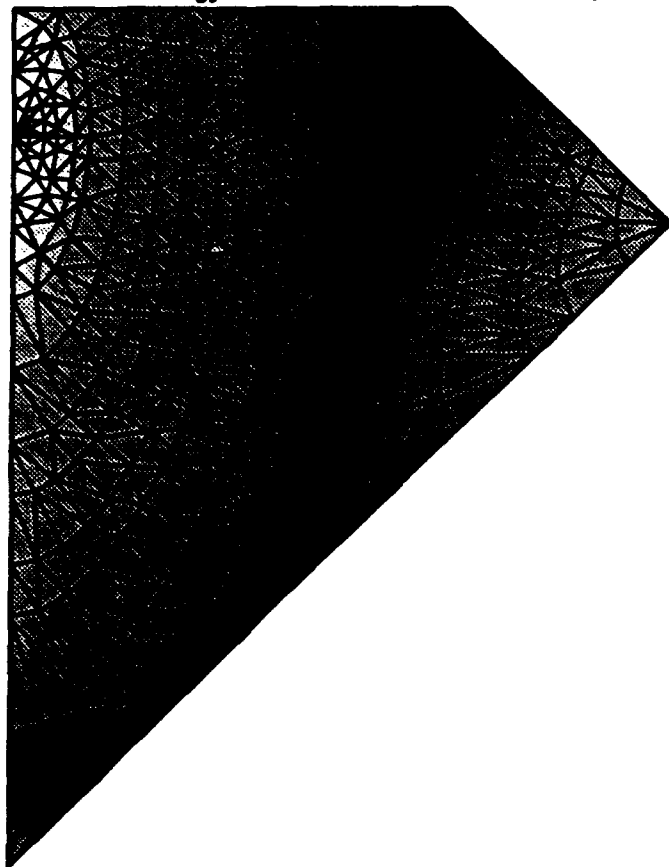
A computationally efficient, full band structure, monte carlo algorithm has been developed. The algorithm is based on a simplex decomposition of phase space, and has all of the advantages of analytic band simulators, while preserving the physics contained in the full band structure. The work required for free-flight computations as well as final state selection is considerably reduced through the use of piecewise linear approximations to both the energy and the electrostatic potential. This method treats motion in k -space and x -space symmetrically allowing exact integrations of the equations of motion and final state selection.

In k -space, the irreducible wedge is represented by a contour-aligned tetrahedral grid. An adaptive grid algorithm is used to generate a mesh that approximates the energy bands to a pre-defined accuracy. Furthermore, the equal-energy surfaces may be readily calculated providing an accurate final k -state selection. As a result, the simulation error is reduced to the monte carlo statistical error and a consistently bounded discretization error deriving from the choice of grid in phase-space.

References

- [1] R. E. BANK AND R. K. SMITH, *A posteriori error estimates based on hierarchical bases*, SIAM J. Numer. Anal., (1993).
- [2] J. BUDE, E. GROSSE, AND R. K. SMITH, *Highly efficient phase-space simplex full band monte carlo simulations*. to be published.
- [3] J. BUDE AND R. K. SMITH, *Phase-space simplex monte carlo for semiconductor transport*, Proceedings of HCIS-8, (1993).
- [4] A. CAYLEY, *On contour and slope lines*, Philosophical Magazine, XVIII (1859), pp. 264-268.
- [5] J. R. CHELIKOWSKY AND M. L. COHEN, *Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors*, Phys. Rev. B, 14 (1976), pp. 556-582.
- [6] M. V. FISCHETTI AND S. E. LAUX, *Full band monte carlo program for electrons in silicon*, Phys. Rev B, 38 (1988), pp. 9721-9745.
- [7] —, *Monte carlo analysis in semiconductor devices: the DAMOCLES program*, IBM J. Res. Develop., 34 (1990), pp. 444-494.
- [8] C. JOHNSON AND E. H. GROSSE, *Interpolation polynomials, minimal spanning trees and ridge-line analysis in density map interpretation*, American Crystallographic Association Program and Abstracts, 4:2 (1976), p. 48.
- [9] S. P. MORSE, *Concepts of use in computer map processing*, Comm. ACM, 12 (1969), pp. 145-152.
- [10] J. L. PFALTZ, *Surface networks*, Geographical Analysis, 8 (1976), pp. 77-93.
- [11] H. SHICHIJO, J. Y. TANG, J. BUDE, AND D. YODER, *Full band monte carlo program for electrons in silicon*, in Monte Carlo Device Simulation: Full Band and Beyond, K. Hess, ed., 1991.
- [12] F. VENTURI, R. K. SMITH, E. SANGIORGI, M. R. PINTO, AND B. RICCO, *A generalized purpose device simulator coupling poisson and monte carlo transport with applications to deep submicron mosfets*, IEEE Trans. CAD, 8 (1989), pp. 360-369.

Silicon Energy Band 4 Grid ; $z = 0$ plane



1		2		3	
4		5		6	
7		8		9	
10		11		12	
13		14		15	

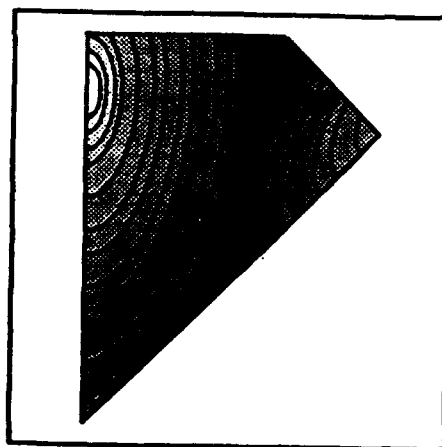


Figure 1: Contour-Aligned Tetrahedral Grid

Ionized Impurity Scattering Model for Monte Carlo Calculations

Jim Dewey and Mohamed A. Osman

**School of Electrical Engineering and Computer Science
Pullman, Washington, 99164-2752**

Abstract

An ionized impurity scattering model for Monte Carlo calculations is proposed. This model includes the following four corrections to the simple Brook-Herring model: 1) momentum dependent screening, 2) the second Born term for single impurity scattering, 3) dressing effect of impurities on electron wavevectors, 4) scattering from pairs of impurities. Monte Carlo calculations are performed for majority electrons in silicon at 300K and 77K with donor concentrations from 10^{15}cm^{-3} to 10^{20}cm^{-3} . Calculations show this model provides a substantial improvement to models based on the first Born approximation.

I. Introduction

Monte Carlo techniques have become a popular method to simulate the characteristics of modern semiconductor devices. In these device simulation programs, an ionized impurity scattering model based on the first Born approximation is commonly used[1,2]. For many materials, such as silicon and GaAs, models based on the first Born approximations overestimate mobility for concentrations greater than 10^{16}cm^{-3} at room temperature[3]. With ionized impurity concentrations in many semiconductor devices greater than 10^{16}cm^{-3} , an improved impurity scattering model for Monte Carlo calculations is clearly needed. Recently Kay and Tang[4] proposed an improved model for Monte Carlo calculations. Their model included phase-shift analysis and multiple impurity scattering where they calibrated their model to fit experimental data. Fischetti[5] showed excellent quantitative agreement with experimental data for both *n* and *p* doped silicon at 300K by including plasmon scattering to phase-shift analysis and short-range carrier-carrier scattering. Another approach which is becoming more popular when simulating coulombic interactions is Molecular Dynamics. Ferry and workers[6] obtained excellent agreement with experimental results for GaAs at 77K. Molecular Dynamics has the disadvantage of requiring large amounts of CPU time.

In this work we include four corrections to the simple Brooks-Herring impurity scattering model and incorporate these in an Ensemble Monte Carlo program. These corrections include momentum dependent screening, dressing effects on the electron wavevectors, the second Born approximation, and multiple-impurity scattering. Mobility is calculated for majority electrons in silicon at 300K and 77K for impurity concentrations ranging from 10^{15}cm^{-3} to 10^{20}cm^{-3} .

II. Electron Transport Model

The electron transport model contains all six non-parabolic ellipsoidal x valleys. Three intervalley g -type phonons between parallel valleys and three intravalley f -type between perpendicular valleys, elastic acoustic phonon scattering, electron-electron scattering and impurity scattering are considered. The phonon scattering and material parameters given by Brunetti and coworkers[7] and the phonon scattering rates calculated by Jacoboni and Reggiani[8] are used in this work. Electron-electron scattering rates take into account the ellipsoidal nature of the conduction band as developed by Osman and coworkers[9]. The impurity scattering model is discussed in the next section.

III. Electron-Impurity Scattering Model

A review of the procedure used to implement corrections to the basic Brooks-Herring impurity scattering model is given in this section. A more complete description will be published elsewhere[10].

As discussed in the introduction, four corrections to the standard Brooks-Herring impurity scattering model are implemented in our Monte Carlo program. The standard Debye screening length is modified by including a momentum dependent correction using the equation developed by Chung and Ferry[11]. Momentum dependent screening results in a scattering rate which is dependent upon the scattering angle θ . For a screened coulombic potential based on the first Born approximation the scattering rate becomes

$$S(k, \theta) = \Omega(k) \frac{1}{\beta^2(\theta) \left[\frac{\beta^2(\theta)}{2k^2} + 1 - \cos \theta \right]^2} \quad (1)$$

where $\Omega(k)$ is the wavevector dependent constant normally seen in the Brooks-Herring scattering rate equation. This scattering rate cannot be calculated in its present form, since the state after scattering is not known prior to the scattering event. This is dealt with in Monte Carlo calculations by using a large correction to the inverse screening length resulting in a scattering rate with internal self-scattering. When impurity scattering is chosen, the scattering angle is selected using a rejection method. Once the scattering angle is known, it is then determined whether impurity or self-scattering terminated the particle's free flight.

The three other corrections are developed following a procedure similar to Moore[12] which is also described by Langer[13] and Rickaysen[14]. The basic procedure is to expand the self energy and define a dressing function as the real part and a width function as the imaginary part of the self-energy terms[15]. In this work, we included the first two terms of the self energy for single impurity scattering which can be thought of as the first and second Born terms and the first term for scattering by a pair of impurities. The dressing effect on the electron wavevectors is included only on the first single impurity term. The dressing effect on electron energy was small compared to other the corrections, this was also observed by Moore[16], and is not included.

When the dressing effect of the impurities on the electron's wavevector is included, the

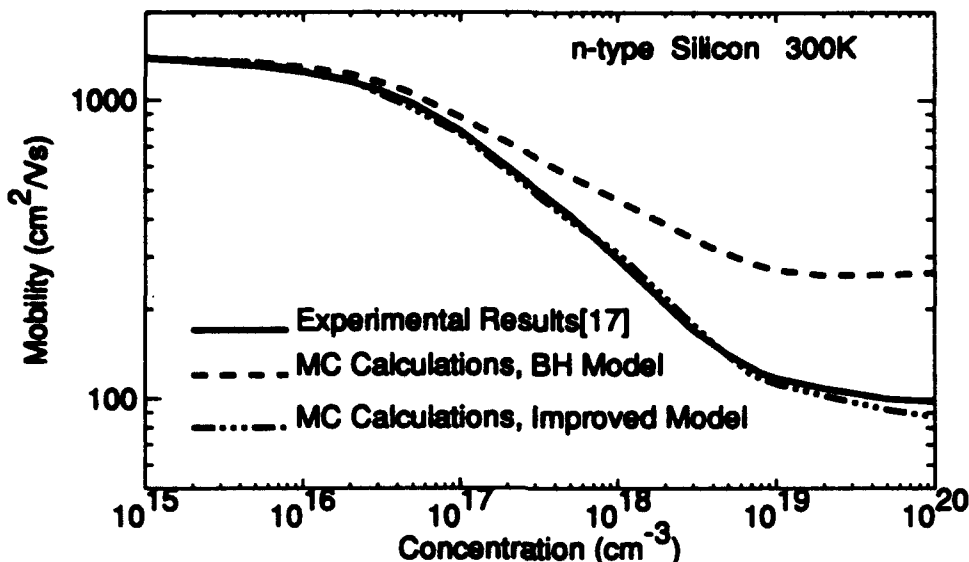


Figure 1: Majority Electron Mobility as a function of ionized impurity concentration for silicon at 300K

scattering rate based on the first Born term becomes

$$S_{\text{screening}}(k) = S_{\text{bh}} \left[1 - \frac{nm^2e^4}{8\pi\hbar^4\epsilon^2\beta^3} \frac{(3\beta^2 + 4k^2)}{\beta^2 + 4k^2} \right]^{-2} \quad (2)$$

where S_{bh} is the scattering rate using the Brooks-Herring model, n is the impurity concentration, m is the electron mass, e is electron charge, ϵ is the material constant, and \hbar is Planck's constant. If momentum dependent screening is used, S_{bh} is given by equation (1) and is incorporated in the same manner.

The scattering rate for the second Born term is given by

$$S(k, \theta) = \frac{nm^2e^6}{4\pi^2\hbar^5\epsilon^3\beta^4} \left[\frac{\tan^{-1} \left(\frac{x}{1+y+4x} \right)^{1/2}}{\sin(\theta/2)(1+y+4x)(1+4x)} \right] \quad (3)$$

where $x = \frac{\hbar^2}{\beta^2} \sin^2(\theta/2)$ and $y = 4k^2/\beta^2$. Equation (3) is integrated numerically over θ to obtain the scattering rate and a rejection method is used to determine the scattering angle θ . As observed by other researchers[16], the correction to mobility by the second Born term is large in region where the first Born term alone is valid $k^2/\beta^2 \gg 1$. This is dealt with by setting the scattering rate for the second Born term to zero when the first Born term alone is valid.

The multi-particle transition rate is found by considering only the lowest approximation for scattering from a pair of impurities. This is the same form as developed by Moore[12] and Rickayson[14]. The resulting scattering rate is quite complex and will not be given

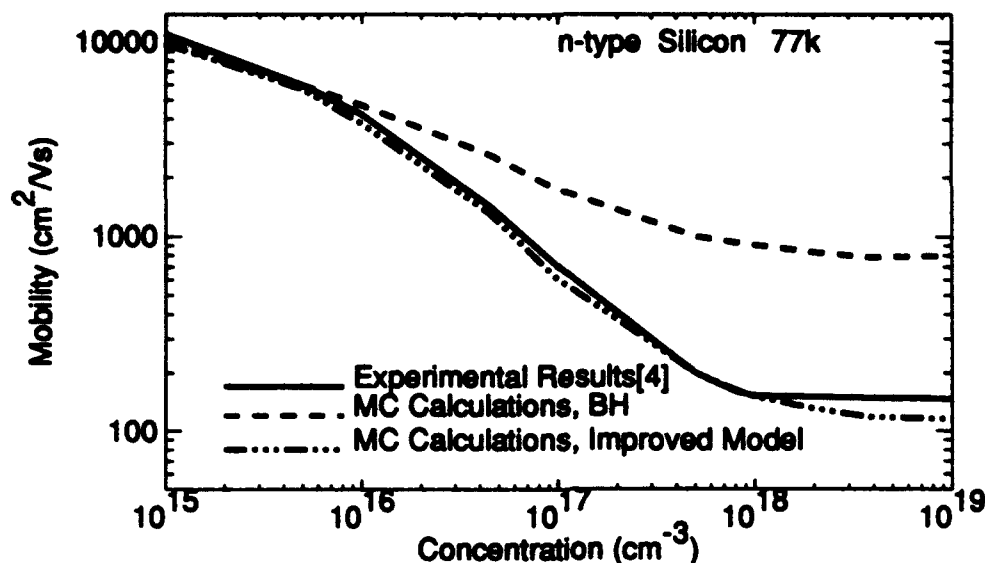


Figure 2: Majority Electron Mobility as a function of ionized impurity concentration for silicon at 300K

here. Multiple impurity scattering is incorporated in the same manner as the second Born term.

IV. Results

Figures 1 and 2 show a comparison of the basic Brooks-Herring model, the model discussed in this paper, and experimental data[17] for silicon at 300K and 77K, respectively. As figure 1 shows, the proposed model clearly improves upon the Brooks-Herring impurity scattering model and compares favorably with experimental data over the range of concentrations considered. Figure 2 shows the discrepancy between the Brook-Herring model and experimental data widens as the temperature is decreased. Again, the proposed model clearly improves upon the Brooks-Herring model.

V. Conclusion

We have proposed an improved impurity scattering model for Monte Carlo calculations. The improved model includes four corrections to the Brooks-Herring impurity scattering model. Results for *n*-doped silicon at 300K and 77K show our model substantially improves mobility calculations compared to models based on the first Born approximation.

VI. References

- [1] G.U.Jenson, G.Lund, T.A.Fjeldly, and M.Shur, "Monte Carlo simulation of semiconductor devices," Computer Physics Comm., vol. 67, pp.1-61, 1991.
- [2] M.V.Fischetti and S.E.Laux, "Monte Carlo analysis of electron transport in small

- semiconductor devices including band-structure and space-charge effects," *Phys. Rev B*, vol. 38, pp.9721-9745, 1988.
- [3] D.Chattopadhyay and H.J.Queisser, "Electron Scattering by ionized impurities in semiconductors," *Rev. Mod. Phys.*, vol. 53, pp.745-768, 1981.
 - [4] L.E.Kay and T.W.Tang, "An Improved ionized-impurity model for Monte Carlo simulations," *J. Appl. Phys.*, vol. 70, pp.1475-1482, 1991.
 - [5] M.V.Fischetti, "Effect of the electron-plasmon interaction on electron mobility in silicon," *Phys. Rev B*, vol. 44, pp.5527-5534, 1991.
 - [6] D.K.Ferry, A.M.Kriman, M.J.Kann, and R.P.Joshi, "Molecular dynamics extensions of Monte Carlo simulation in semiconductor modeling," *Computer Physics Comm.*, vol. 67, pp.119-134, 1991.
 - [7] R.Brunetti, C.Jacoboni, F.Nava, and L.Reggiani, "Diffusion Coefficient of Electrons in Silicon," *J. Appl. Phys.*, vol. 62, pp.6713-6722, 1981.
 - [8] C.Jacoboni and L.Reggiani, "Monte Carlo Method in Transport," *Rev. Mod. Phys.*, vol. 55, pp. 645-705. 1983.
 - [9] M.A.Osman, N.Nintunse, and M.Imam, "Carrier-carrier interaction effects on transient valley repopulation and velocity in silicon," *Semicond. Sci. Technol.*, vol. 7, B340-B342, 1992.
 - [10] J.Dewey and M.A.Osman, "Monte Carlo Calculation of Electron Mobility in Silicon," (in preparation).
 - [11] W.Chung and D.K.Ferry, "Dynamic Screening for Ionized Impurity Scattering in Degenerate Semiconductors," *Solid-State Electronics*, vol. 31, pp.1369-1374, 1988.
 - [12] E.J.Moore, "Quantum-Transport Theories and Multiple Scattering in Doped Semiconductors. I. Formal Theory," *Phys. Rev. B*, vol. 160, pp. 607-616, 1967.
 - [13] J.S.Langer, "Theory of Impurity Resistance in Metals," *Phys. Rev. B*, vol. 120, pp.714-725, 1960.
 - [14] G.Rickayson, "Scattering by Impurities in Metals," *Lecture Notes on the Many-Body Problem*, edited by C. Fronsda, 1961.
 - [15] L.P.Kadanoff and G.Baym, "Quantum Statistical Mechanics," *Frontiers in Physics*, edited by D.Pines, 1962.
 - [16] E.J.Moore, "Quantum-Transport Theories and Multiple Scattering in Doped Semiconductors. II. Mobility on *n*-type GaAs," *Phys. Rev. B*, vol. 160, pp. 618-626, 1967.
 - [17] G.Masetti, M.Severi, and S.Solmi, "Modeling of Carrier Mobility Against Carrier Concentration in Arsenic-, Phosphorus-, and Boron-Doped Silicon," *IEEE Trans. Electron Devices*, vol. ED-30, pp.764-769, 1983.

A Monte Carlo Method for the Study of Non-equilibrium Phenomena in Low Dimensional Semiconductor Structures

L. Rota ^{a)}, F. Rossi ^{b)}, P. Lugli ^{c)}, and E. Molinari ^{d)}

- a) Department of Physics, Clarendon Laboratory
University of Oxford, Parks Road, Oxford OX1 3PU, UK
- b) Philipps-Universität Marburg, Fachbereich Physik und Zentrum
für Materialwissenschaften, Renthof 5, D-3550 Marburg, Germany
- c) Dipartimento di Ingegneria Elettronica, II Università di Roma
Via della Ricerca Scientifica, 00133 Roma, Italy
- d) Dipartimento di Fisica, Università di Modena
Via Campi 213/A, 41100 Modena, Italy

Abstract

We present a numerical study of the non-equilibrium properties of quasi-one dimensional systems with different confining geometries. Carrier cooling following a laser photoexcitation is analyzed by means of a Monte Carlo simulation which includes all relevant scattering mechanisms. Our results indicate that, for wire made available by today's technology, the carrier dynamical properties of these systems are not so different with respect to bulk GaAs.

I. Introduction

The recent advances in compound semiconductor technology have allowed the fabrication of wire-like regions where quasi 1D confinement is achieved in a narrow-band-gap semiconductor material surrounded by another large-band-gap semiconductor [1]. In particular, such wire-like structures have been fabricated with rectangular and V shaped cross sections whose dimensions are comparable to the carrier coherence length [2,3]. In such structures, carrier transport characteristics are expected to differ from the bulk case as the carrier-phonon scattering rate is affected by the changes in the electronic as well as in the vibrational properties induced by layering [4].

Much of the theoretical work dedicated so far to one dimensional structures has covered mainly the so called "extreme quantum limit", i.e. the case in which only one subband is occupied [1,5]. For wire structures available today, such limit can only be reached under extremely low temperature and electric field conditions. Usually, several subbands are present, and their contribution to carrier transport has to be taken into account.

The aim of the present contribution is to extend the analysis of charge transport in quasi one-dimensional systems to nonequilibrium conditions and nonlinear regimes, by studying realistic wire-like structures where many subbands are normally occupied. This is performed by means of an Ensemble Monte Carlo simulation which allows the study of dynamical evolution of photoexcited carriers, a situation often achieved, for instance, in ultrafast optical experiments. By using the electronic levels and wave functions as derived from an exact solution of Schrödinger's equation, we evaluate the electron-phonon and electron-electron scattering rates which are, in turn, used in the simulation to study the carrier relaxation process. In addition, the method allows us to easily include effects like degeneracy and non-equilibrium phonons.

We will investigate two different type of GaAs-based structures, namely rectangular wires obtained from multi quantum well samples by chemical etching [2,6] and V-grooved wires obtained by growing a thin quantum well on an etched V-shaped profile [7].

II. Electronic states and electron-phonon interaction

Chemical etching gives GaAs-based rectangular quantum wire surrounded along one of the confinement directions by AlGaAs layers, and and free-standing along the other confinement direction [6]. In this case, the conduction band electrons can be reasonably described by a factorized envelope function of the type $\psi(x)\psi(y)e^{ik_z z}$, where $\psi(x)$ and $\psi(y)$ are respectively the solution of the one dimensional Schrödinger equation in a rectangular potential.

The situation is more complicated for the V-grooved structure. There, in order to study the energy levels and associated eigenfunctions of a two-dimensional V-like profile, we consider the following single-particle Schrödinger equation

$$\left(-\frac{\hbar^2 \nabla^2}{2m} + V(x, y) \right) \psi(x, y) = \mathcal{E} \psi(x, y) \quad (1)$$

where ∇^2 denotes the two-dimensional Laplace operator and $V(x, y)$ the two-dimensional potential profile. By considering the following set of two-dimensional planewaves over a rectangular domain Ω :

$$\phi(\mathbf{k}_x, \mathbf{k}_y; x, y) = \frac{1}{\sqrt{\Omega}} e^{i(\mathbf{k}_x x + \mathbf{k}_y y)}, \quad (2)$$

the Schrödinger equation (1) can be rewritten as

$$\left(\frac{\hbar^2 (\mathbf{k}_x^2 + \mathbf{k}_y^2)}{2m} + V(\mathbf{k}_x, \mathbf{k}_y; \mathbf{k}'_x, \mathbf{k}'_y) \right) c(\mathbf{k}'_x, \mathbf{k}'_y) = \mathcal{E}(\mathbf{k}_x, \mathbf{k}_y), \quad (3)$$

where the coefficients $c(\mathbf{k}'_x, \mathbf{k}'_y)$ are the Fourier component of the total wavefunction $\psi(x, y)$ and $V(\mathbf{k}_x, \mathbf{k}_y; \mathbf{k}'_x, \mathbf{k}'_y)$ are the matrix elements of the potential profile in the plane-wave basis given in Eq. (2). By means of a standard numerical procedure, we derive the eigenvalues \mathcal{E} corresponding to the energy levels and the Fourier coefficients $c(\mathbf{k}'_x, \mathbf{k}'_y)$ of the corresponding eigenfunction. Once such coefficients are known, we obtain for each energy level \mathcal{E} the corresponding eigenfunction according to $\psi(x, y) = \sum_{\mathbf{k}'_x, \mathbf{k}'_y} c(\mathbf{k}'_x, \mathbf{k}'_y) \phi(x, y)$. Our numerical results have been obtained for the case of a V-like potential profile shown in Fig. 1. The V-like region is characterized by a potential V_1 and all the surrounding region by a potential V_2 . Fig. 2 shows a plot of the carrier density corresponding to the first four electronic levels. As expected, the ground state exhibits a single maximum in the center of the V-like region, while the excited states extend over the whole V-groove exhibit an increasing number of maxima as the order of the level increases.

Starting from the electronic wave functions and energy levels just described, we have computed the electron-phonon scattering rate for the two structures, assuming in both cases bulk LO phonon modes. This approximation should be quite acceptable in these

Fig. 1. Section of the V-grooved quantum wire.

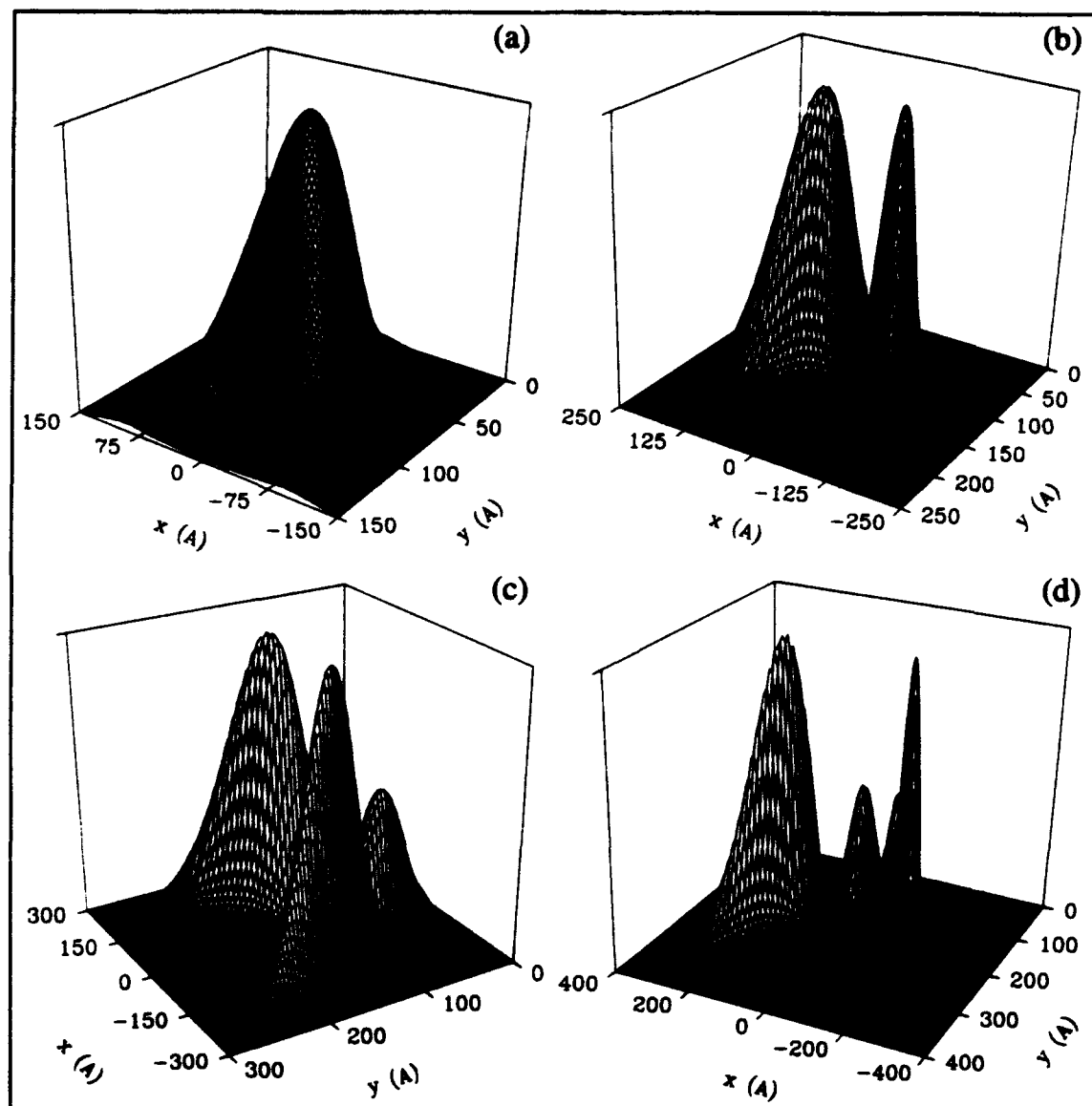
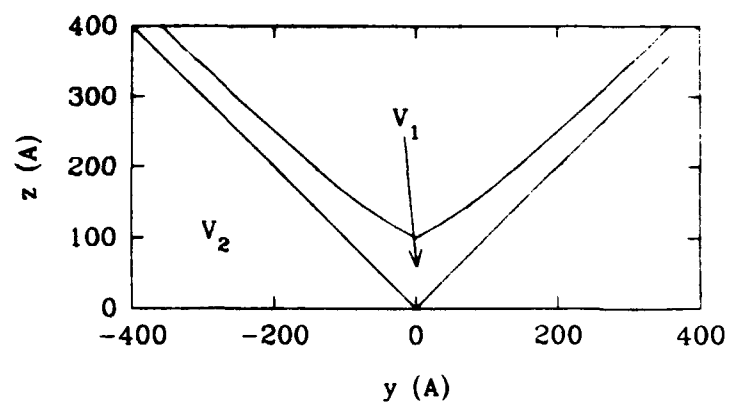


Fig. 2. Electron charge density in the V-grooved quantum wires for the first four energy levels.

systems, which have a large spatial extension, due to the recent findings that the overall electron-LO phonon scattering rate in quantum wells does not differ significantly from the bulk one once the contribution of confined and interface phonon modes is properly accounted for [8]. Figures 3 (a) and (b) show respectively the scattering rate for the rectangular quantum wire and the V-grooved wire (solid lines) [9], compared with bulk GaAs rate (dashed line). In the V-grooved structure, only the first five energy levels are accounted for. Apart from the different position of the energy levels in the two quantum wires, the behavior is very similar and the overall scattering rates in both wires is similar to that of the bulk.

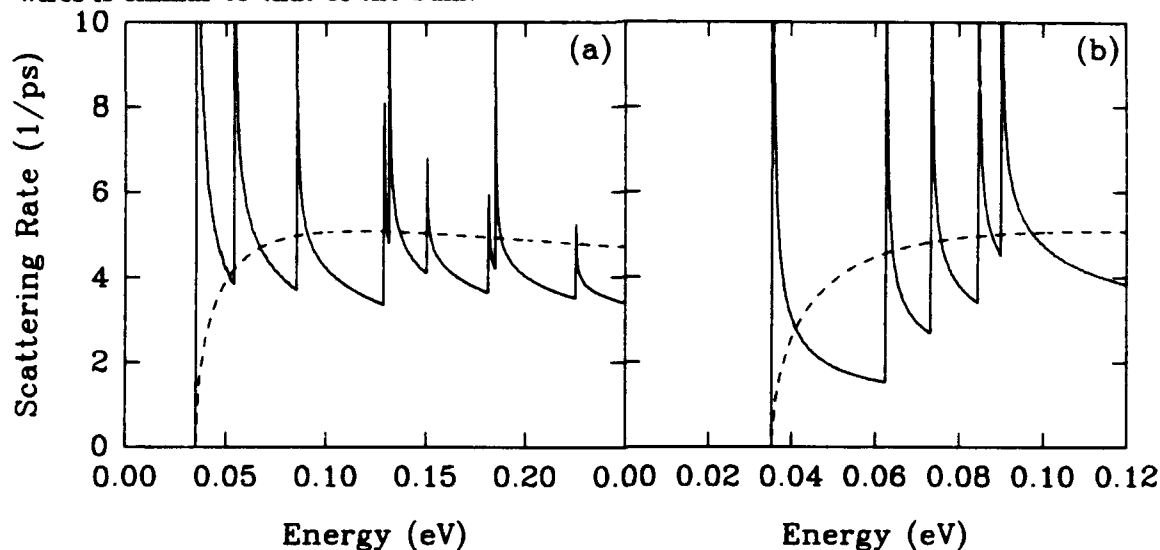


Fig. 3. Electron-phonon scattering rate for (a) a rectangular cross section quantum wires and (b) a V-grooved quantum wires (solid lines), compared with the bulk scattering rate (dashed line).

III. Simulation results

As the scattering rate for the V-groove and the rectangular wire are similar we performed the simulation only for the rectangular wire. We expect that the results will be qualitatively the same for the V-groove wire. Our Monte Carlo simulation accounts for the electron-polar optical phonon as well as the electron-electron interaction for both intrasubband and intersubband transitions. The model used here is similar to that described in ref. 10 but the effect of non-equilibrium phonons have been accounted for with a rejection technique. The study of the effect played by the hot phonon population is more difficult in this system, compared to the bulk case, due to the lack of isotropy. The phonon population has been collected using a three-dimensional mesh. The cooling of the excited carriers following a laser excitation is shown in Fig. 4 (a), (b) and (c) respectively for the one dimensional densities of 10^5 cm^{-1} , $3 \times 10^5 \text{ cm}^{-1}$ and 10^6 cm^{-1} . As already verified [11] the cooling in the quantum wire (solid line) is slower compared to the bulk (dashed line) due to the reduced intersubband scattering rate, but this effect is modified by the presence of the hot phonon population. At very low densities, i.e. when the phonon population is not driven out of equilibrium, there is a big difference in the cooling between the wire and the bulk. Vice versa at the higher density this difference is strongly reduced. This effect can be clearly attributed to the

reduced efficiency of hot phonon reabsorption in the quantum wire caused by the spatial anisotropy of the system.

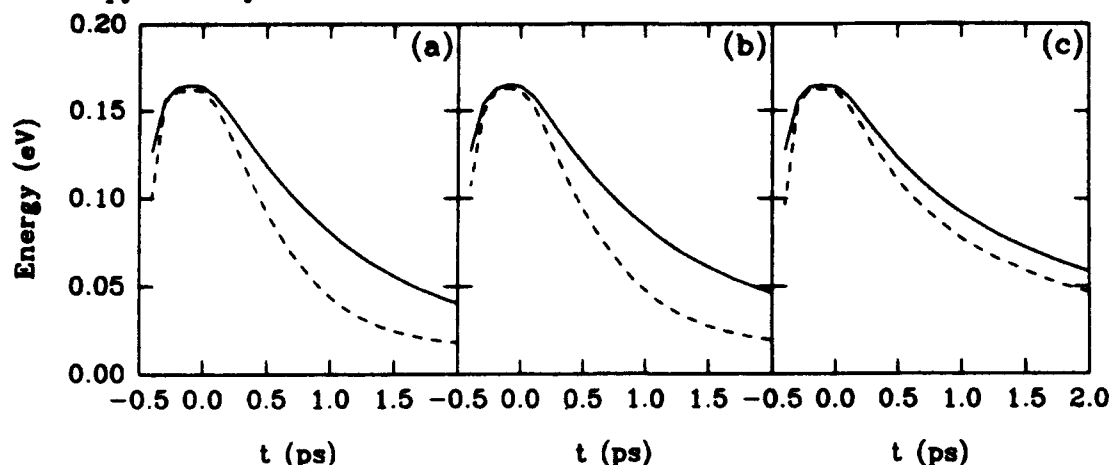


Fig. 4. Electron mean energy after a laser photoexcitation in a quantum wire (solid lines) for three different densities: (a) 10^5 cm^{-1} , (b) $3 \times 10^5 \text{ cm}^{-1}$ and (c) 10^6 cm^{-1} , compared with the equivalent results in bulk GaAs (dashed lines).

In conclusion, we have shown that the Monte Carlo method can be easily applied to the study of low dimensional systems described by any type of geometrical potential confinement. The dynamical properties of the carriers can be investigated in strongly non-linear and non-equilibrium conditions with any number of occupied subbands. Our results show that, for the relatively large wire made available by the actual technology, the difference between the behaviour of these system and a bulk system is much lower of that predicted and probably only in systems extremely narrow a faster carrier transport will be found.

References

- [1] *Nanostructure Physics and Fabrication*, ed. by M.A. Reed and P. Kirk, Accademic Press, Boston, 1989.
- [2] T. Demel, D. Heitmann, P. Granbow, and K. Ploog, *Phys. Rev. B* **38**, 12732 (1988);
- [3] E. Kapon, M. C. Tamargo, and D. M. Hwang, *Appl. Phys. Lett.*, **50**, 347 (1987).
- [4] U. Bockelmann and G. Bastard, *Phys. Rev. B* **42**, 8947 (1990).
- [5] A. Kumar and P.F. Bagwell, *Phys.Rev. B* **43**, 9012 (1991).
- [6] R. Cingolani, H. Lage, L. Tapfer, H. Kalt, D. Heitmann, and K. Ploog, *Phys. Rev. Lett.* **67**, 891 (1991).
- [7] U. Marti, M. Proctor, R. Monnard, D. Martin, F. Morier-Gemoud, F. K. Reinhart, R. Widmer, H. Lehmann, *APC91, American Vacuum Society Conference Proceedings*, **227**, 80 (1991).
- [8] H. Ruecher, E. Molinari, and P. Lugli, *Phys. Rev. B* **45**, 6747 (1992).
- [9] L. Rota, F. Rossi, P. Lugli, E. Molinari, S. M. Goodnick, and W. Porod, *SPIE*, Vol. 1676, 161 (1992).
- [10] L. Rota, F. Rossi, S. M. Goodnick, P. Lugli, E. Molinari, and W. Porod, *Phys. Rev. B* **47**, 1632 (1993).
- [11] P. Lugli, L. Rota, and F. Rossi, *Phys. Stat. Sol. (b)*, **173**, 229 (1992).

Monte Carlo Modelling of Sub-Micron Delta-Doped MOSFETs

R W Kelsall and A G O'Neill

Department of Electrical and Electronic Engineering,
University of Newcastle upon Tyne,
Newcastle
NE1 7RU, UK

Abstract

A 2D multilayer MOSFET simulator has been developed, using self-consistently coupled ensemble Monte Carlo and 2D Finite Element Poisson Solver algorithms. The simulator is used to investigate the operation of sub-micron delta-doped MOSFETs, in order to assess their suitability for high density logic applications. An operating window is observed within which the delta-doped devices exhibit a significant reduction in surface carrier density, relative to conventional MOSFETs, implying reduced surface scattering and gate injection.

1. Introduction

In order to achieve Ultra Large Scale Integration (ULSI) densities in CMOS, the gate lengths of silicon MOSFETs must be reduced well below $1\mu\text{m}$. It is well known that sub-micron MOSFETs are prone to a number of short channel effects: increased lateral electric fields result in enhanced hot carrier generation, whilst increased vertical fields pull carriers hard onto the Si/SiO₂ interface - leading to increased interface scattering and hot carrier injection into the SiO₂ layer. The latter effect results in an accumulation of trapped charge in the oxide, which causes eventual device failure by shifting the threshold voltage and reducing the ability of the gate to modulate charge.

These problems may be alleviated by using a delta doped MOSFET structure [1,2], in which conduction occurs, not at the Si/SiO₂ surface, but in an ultrathin, highly doped layer located typically 200-400Å below the interface. Thus, both interface scattering and hot carrier injection should be reduced: however, realistic modelling work is necessary to determine the efficacy of the delta layer in confining carriers, its effect on the surface carrier density and energy distribution, and the extent of any consequent reduction in hot carrier degradation.

In this paper, we describe the development of a 2D self-consistent Monte Carlo simulator, and its application to modelling electronic transport in delta-doped MOSFETs. In very short gate devices transport can no longer be described in terms of steady state mobilities and saturation velocities, and the traditional drift-diffusion approach to device modelling becomes inadequate. The Monte Carlo method is well established as a powerful technique for modelling highly non-equilibrium transport phenomena, and can yield a detailed microscopic insight into the operation of semiconductor devices. The initial investigation presented here concerns the basic issues of transistor operation and carrier distribution in short gate delta doped MOSFETs, and the effect of the delta-doped layer on the surface electron density.

II. The Simulation Program

We have developed a high specification 2D simulation program for sub-micron multilayer MOSFETs. The program consists of an ensemble Monte Carlo algorithm self-consistently coupled to a 2D Finite element Poisson Solver. The 2D Poisson equation is re-solved every 1fs: such a short timestep is necessary to avoid undersampling of the plasma oscillations which occur in the highly doped source and drain implants of silicon MOSFETs.

The Finite Element mesh can be defined by the user, but is restricted to a rectangular grid format. This restriction greatly simplifies the determination of the 2D charge density in the device from the distribution of Monte Carlo particles. The ability to specify non-uniform mesh spacings is essential for both delta-doped and conventional MOSFETs, in order that the potential at, in the former case, the delta-doped layer, and, in the latter case, the surface inversion layer, may be accurately modelled.

Electronic transport in the MOSFETs is modelled using a nonparabolic, ellipsoidal description of the 6 silicon X-valleys. Intervalley, and acoustic intravalley phonon scattering is included using the parameter set recommended by Brunetti et al.[3], which was found to produce good agreement with experimental drift velocity data. Impurity scattering is included via the Brooks-Herring model.

The effect of degeneracy (Pauli exclusion) is modelled using the approach described by Lugli and Ferry [4], in which the probability that an electron is scattered into a state of energy E is weighted by $1-f(E)$, where $f(E)$ is the local electronic distribution function. In our simulation, the distribution function may be sampled in up to 8 different spatial regions - including the source and drain implants, and sections along user defined surface and buried layers - to account for spatial variations in electron density and carrier heating.

In order to model self-consistently the electrostatic potential at the pn junctions between the implants and the channel region, a 2D hole density is included in a zero current (fixed quasi Fermi level) approximation, as described by Fischetti and Laux [5]. A damping scheme is employed to ensure a smooth convergence to the fixed quasi Fermi level condition. The effect of the hole density on the electrostatic potential is particularly important for very short gate devices, where the acceptor density must be increased in order to combat drain induced barrier lowering and punchthrough.

A particle replication scheme [6] is utilised to increase the sampling capability of the simulator in the channel region of the devices. For a replication factor M , any particle entering the channel region is replicated $M-1$ times. Particles attempting to enter an implant region are annihilated with a probability $(M-1)/M$. Correspondingly, all particles in the implant regions have an associated charge M times greater than those in the channel region.

The source-drain current is calculated by summing the x-component of velocity for all particles located between the source and drain implants [7]. This method enables a much larger sample to be used than in the traditional approach of counting particles emitted/injected at the source and drain contacts.

Several authors have modelled surface scattering in MOSFETs by allowing both specular and diffusive reflection of particles which impinge on the oxide interface [8]. Our program allows any proportion of diffusive reflections to be specified; however, for this particular investigation, we have used specular reflections only. The extent of surface scattering is primarily important in determining the relative speed of delta-doped and conventional MOSFETs, an issue which will be the subject of a separate investigation. For the same reason we have not, at this stage, included any sophisticated description of impurity scattering in the delta-doped layer: such refinements may readily be made, as and when required.

The program runs for typically 2-5 hours per bias point on an HP 710 workstation. The actual run time depends on the device size and bias conditions, with simulations of shorter devices generally taking less time. Access to a large networked cluster of workstations at Newcastle means that a simulations for a whole set of bias points can be run in parallel, on the same timescale.

III. Results

We have modelled a $0.1\mu\text{m}$ delta-doped MOSFET with the following specification. The delta doping dose was 10^{12}cm^{-2} - implemented as a 20\AA doping plane, with a bulk doping density of $5 \times 10^{18}\text{cm}^{-3}$. The delta-doped layer was located 200\AA below the Si/SiO_2 interface. A doping density of $2 \times 10^{19}\text{cm}^{-3}$ was used for the source and drain implants: whilst this value is lower than those encountered in some MOSFET structures, it was found sufficient to define a flat potential across the implant region - hence providing a suitable model of an ohmic contact. The implant depth was taken as 500\AA , and a length of 1000\AA of each implant was included in the simulation. Again, this length was found sufficient to model the ohmic contact and the surrounding potential. The use of higher implant doping densities, or the inclusion of a greater length of implant, leads to large increases in CPU time, with no significant gain in physical information or quantitative accuracy.

Short gate MOSFETs can suffer from high leakage currents due to drain induced barrier lowering. The effect can be alleviated by increasing the substrate doping and, in common with previous reports on $0.1\mu\text{m}$ conventional MOSFETs, we have specified a doping density of 10^{17}cm^{-3} for the p-type silicon substrate. Threshold voltage shifts are also a problem in short gate FETs: we have chosen an oxide thickness of 50\AA , in order to reduce the gate voltage swing required to turn the device off. Assuming an n-type polysilicon ohmic contact, our simulations predict a threshold voltage of around -1.5V for the device under investigation. Obviously, the threshold voltage varies with the delta layer depth and dose. We have used a drain bias of 1V in our simulations, reflecting the reduced supply voltage anticipated for use in ULSI.

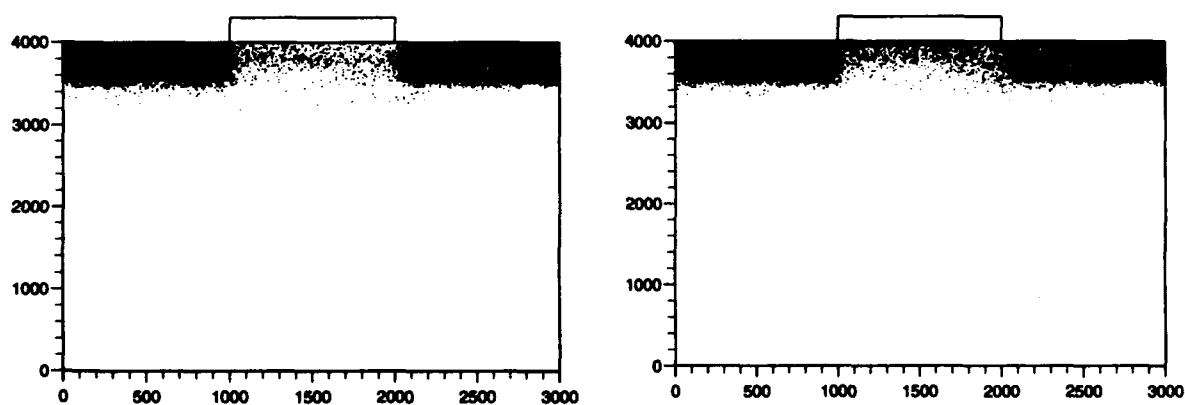


Figure 1. Steady state electron distribution: (a) $V_G=0\text{V}$, $V_D=1\text{V}$ (b) $V_G=0.75\text{V}$, $V_D=1\text{V}$

Figures 1(a) and (b) show instantaneous plots of the 2D electron distribution in the device, during steady state operation, for gate biases of 0V and 0.75V. Figure 1(a) shows conduction occurring well below the oxide interface; however, it is clear that the delta layer does not provide strong carrier confinement, with electrons travelling in a very broad current channel. On the other hand, in figure 1(b), for $V_G = 0.75V$, an inversion layer has formed at the interface; the device is now operating primarily in a surface channel mode. This type of transition has also been observed in drift-diffusion modelling of long channel devices [9].

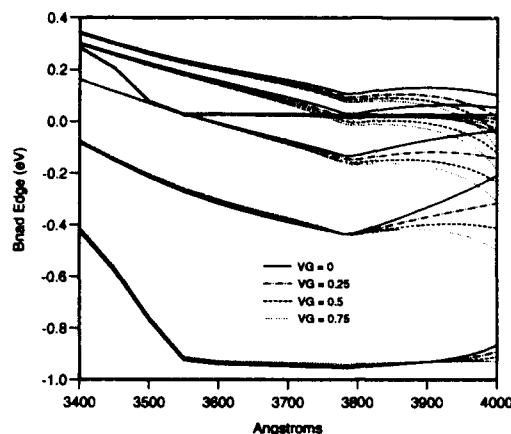


Figure 2. Conduction band edge profile vs. vertical distance, with 4000Å corresponding to the oxide interface. For each gate bias, a set of six 'slices' of the potential profile are shown, taken at 200Å intervals along the channel from source to drain. The profiles corresponding to the endpoints of the channel are pinned close to the source and drain biases for most of the vertical distance shown, whilst the other four profiles appear in descending order moving along the channel from source to drain.

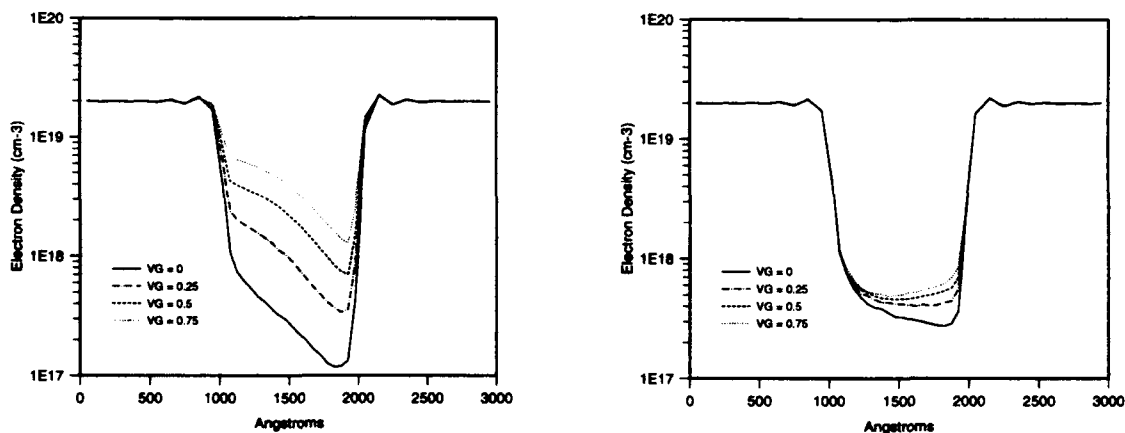


Figure 3. Electron density vs. distance from source to drain: (a) at the oxide interface, (b) along the delta-layer

Figure 2 shows several sets of vertical 1D slices through the conduction band edge profile, for a range of gate biases. In this figure it can be seen that the potential 'notch' caused by the

delta layer is quite shallow and, as the gate bias is increased, the potential at the oxide interface drops below that in the delta layer, along most of the channel length. Figures 3(a) and (b) show the electron density along the surface layer and the delta layer, for the same range of gate biases. Comparing the two figures, it is clear that the surface electron density exceeds the delta layer density for all the gate biases shown except for $V_G = 0$. This conclusion is consistent with the potential profiles shown in figure 2, which also indicate greater surface accumulation for $V_G > 0V$. At $V_G = 0V$ the drain current for this device is 0.21mA per micron gate width. For comparison, we simulated a 0.1 μ m conventional MOSFET with the same implant, oxide and substrate specification. Approximately the same drain current, 0.21mA/ μ m, was obtained for a gate bias of 0.3V. In figure 4 we have compared the surface electron densities for the two devices. The surface electron density for the delta doped device is less than one third of that in the conventional MOSFET, for the same drain current. An equal current comparison was also performed for the $V_G = 0.25V$ bias condition on the delta doped device, and a reduction in the surface electron density - by a factor of approximately 2.5 compared to the conventional device - was still found, even though the delta-doped MOSFET showed appreciable surface conduction.

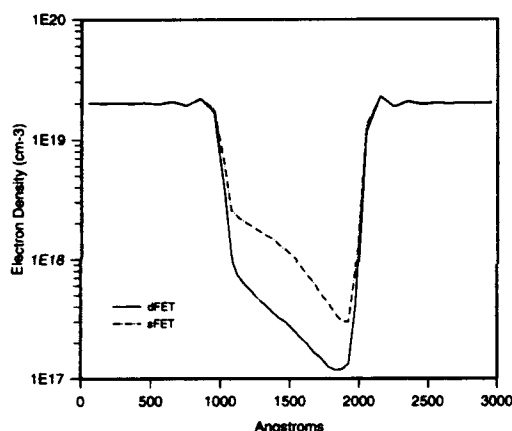


Figure 4. Surface electron density vs. distance from source to drain for a 0.1 μ m delta doped MOSFET (dFET) and a conventional surface channel device (sFET)

Obviously, a different result may be obtained by varying the depth of the delta-layer below the oxide interface. We have carried out simulations of devices with delta-layer depths of 100 \AA , which show that the conducting channel spreads up to the oxide interface, even when no surface inversion layer is formed. Conversely, further reductions in surface carrier density may be achieved by increasing the delta-layer depth, but this reduces the transconductance and the threshold voltage: beyond a certain depth, gate control can be lost altogether [9].

IV. Conclusions

We have used a 2D multilayer MOSFET Monte Carlo simulator to investigate the operation of 0.1 μ m delta-doped MOSFETs. The devices have an operating 'window' of gate biases within which the conducting channel is centred on the delta-doped layer. In this mode, a substantial reduction in the surface carrier density can be obtained, relative to conventional MOSFETs, implying consequent reductions in surface scattering and gate injection. Outside this operating window, a parasitic surface channel is formed, and the above advantages are

diminished. The operating window is large enough to be compatible with logic operation at a reduced supply voltage of 1V.

References

1. A C G Wood, A G O'Neill, P Phillips, R G Biswas, T E Whall and E H C Parker
IEEE Trans. Electron Devices **ED-40** 157 (1993)
2. A G O'Neill, A C G Wood, P Phillips, T E Whall, E H C Parker, A Gundlach and S Taylor
Microelectron. Eng. **19** 743 (1992)
3. R Brunetti, C Jacoboni, F Nava, L Reggiani, G Bosman and R J J Zijlstra
J. Appl Phys. **52** 6713 (1981)
4. P Lugli and D K Ferry IEEE Trans. Electron Devices **ED-32** 2431 (1985)
5. M V Fischetti and S E Laux Phys. Rev. B **38** 9721 (1988)
6. A Phillips and P J Price Appl. Phys. Lett. **30** 528 (1977)
7. V Gruzinskis, S Kersulis and A Reklaitis Semicond. Sci. Technol. **6** 602 (1991)
8. P J Price Semiconductors and Semimetals **14** 249 (ed. Willardson and Beer)
9. A C G Wood and A G O'Neill Materials Research Symposium Proceedings
vol **220** (silicon molecular beam epitaxy) 465 (1991)

Techniques for Modeling Ultrasmall Quantum Devices

D. K. Ferry¹ and H. L. Grubin²

¹Center for Solid State Electronics Research
Arizona State University
Tempe, AZ 85287-6206

²Scientific Research Associates, Inc.
P. O. Box 1058
Glastonbury, CT 06033

Abstract

As semiconductor devices continue to shrink in size, quantum effects are beginning to become important in their operation. In this paper, we discuss some approaches for including the quantum effects in device modeling.

I. Introduction

The transport of carriers in semiconductors and in ultra-small semiconductor devices has long been a subject of much interest, not only for material evaluation, but also in the realm of device modeling and, more importantly, as an illuminating tool for delving into the physics governing the interaction of electrons (or holes) with their environment. The scaling of ULSI device dimensions to future chips indicates that we will eventually see devices with gate lengths at the 0.05 μm level. Very few laboratories have produced even working research transistors with gate lengths on the 50 nm scale and little is understood about the limitations (from the physics) that will determine whether or not these devices are practical. On the other hand, when devices of 30 nm (or less) gate lengths are made, it is found that their performance is different from that of current FETs. Research devices with gate lengths of 25-80 nm clearly show that tunneling through the gate depletion region is a major contributor (if not the dominant contributor) to current, and gate control is much reduced due to this effect [1-3]. Even in MOSFETs, quantization is found to occur in the channel, which affects the overall performance of the devices [4]. In consequence, it appears that more detailed modeling of quantum effects needs to be included in device modeling for future ultrasmall devices.

There are several approaches that have been used to model quantum effects in semiconductor devices (of varying levels of device complexity). In this paper, we will try to give a short overview of some of these approaches, and indicate how they are similar and how they differ. In the next section, we discuss how quantum modeling differs from semi-classical modeling. We then turn to a description of the various quantum "distribution" functions, discuss their equations of motion, and the levels of complexity. In each case, simple examples are described where the approach has been used with some effectiveness.

II. How does Quantum Modeling Differ from Classical Modeling?

Generally, modeling of quantum phenomena is more complicated than modeling of classical and/or semi-classical phenomena. For instance, the energy-conserving delta function used in computing scattering rates with the Fermi golden rule is no longer valid, as energy and momentum become separate dynamical variables. Thus, we are forced to add a method of computing the *spectral density*, which relates the energy to the momentum, in addition to having to compute non-equilibrium distribution functions (or various moments of these). To

be sure, in some approaches this is finessed by using single-time functions, such as the density matrix and the Wigner distribution function, which essentially integrate out the spectral function. But, this is accompanied by the full non-local nature of the potential interactions becoming explicit in the dynamical variables; i.e., the potential becomes a two-point function. Let us consider further how this nonlocality arises. Consider a simple potential barrier $V(x) = V_0 u(-x)$, where $u(x)$ is the Heavyside step function. We assume that there is some density existing in the region $x > 0$, and the question is how the density varies near the barrier, a quite typical problem in introductory quantum mechanics, except here we have a statistical mixed state to describe. In Fig. 1, we show the Wigner distribution function for this case (for parameters appropriate to GaAs, with $n = 2 \times 10^{17} \text{ cm}^{-3}$) for $V_0 \rightarrow \infty$. We note that far from the barrier, the distribution approaches the classical Maxwellian form, but near the barrier, it differs greatly. The repulsion from the barrier is required by the vanishing of the wave function at the barrier, but the first peak away from the barrier (in the wave function) occurs closer to the barrier for higher momentum states. This leads to much of the complication evident in the figure. The overshoot occurs to accommodate the need for total charge neutrality. Classically, in the absence of self-consistency, the density would be uniform up to the barrier, and the differences are the result of the quantum mechanics. This variation exists over a distance of the order of several thermal de Broglie wavelengths, which provides a spatial scale length. In GaAs, at 300 K, this is about 5 nm for electrons, and of course increases with the inverse square root of the temperature as the thermal de Broglie wavelength is given by $\lambda_D = (\hbar^2/2mk_B T)^{1/2}$. Thus, nonlocal variations can be expected over a range of 10-20 nm even at room temperature!

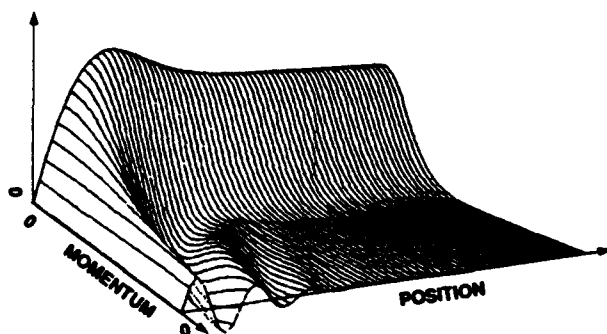


Figure 1. The Wigner distribution function for an infinite barrier, in arbitrary units [5].

It is clear that the density no longer varies simply as $\exp(-\beta V)$, where β is the inverse temperature, and that modifications to the statistical mechanics need to be made. The development of quantum corrections to statistical thermodynamics, especially in equilibrium, has a rich and old history. Unfortunately, there is no consensus as to the form of the quantum potential correction to this simple exponential. If we could find such a correction, it could be utilized in the semi-classical hydrodynamic equations, and most further quantum complications ignored. The various forms of the quantum potential, for use in classical hydrodynamic equations, has recently been reviewed [6]. One form that has been used [7] introduces a *quantum pressure* term as a modification of the electron temperature, through

$$\frac{3}{2} k_B T_{\text{eff}} = \frac{3}{2} k_B T_e - \frac{\hbar^2}{8m^*} \nabla^2 \ln(n), \quad (1)$$

although other work has reduced the last term by a factor of 3 [8,9]. The form of the last term in (1) was originally derived by us and is often termed the Wigner potential [10]. Although the results obtained using this model are in agreement with the intuitive expectations, it should be noted that the correction term is an average and does not have the

momentum dependence expected from Fig. 1. A more recent derivation overcomes this limitation [6], but has not been tested in actual simulations as yet.

Another problem with the use of quantum approaches to device modeling is that most quantum discussions, especially those of quantum transport, tend to revolve around *closed* systems, whereas most devices are *open* systems. In treating such open systems quantum mechanically, it is quite difficult to properly define the reservoir (thermal equilibrium) regions, as well as the contact regions between the reservoirs and the active device region. Because of the nonlocal nature of the quantum system, errors in defining the contact region will propagate throughout the device, often leading to spurious results.

III. Approaches to Quantum Distributions

Why don't we just solve the Schrödinger equation for a given potential distribution, and then weight various solutions with a Fermi-Dirac distribution? This approach actually works well in equilibrium situations, or when we know in detail the properties of all contact regions and the exact potential structure within the active area. However, finding the steady-state solutions with the above approach often entails more work than using one of the techniques for directly finding a quantum distribution self-consistently within the entire device domain. A more important reason for not a simple Schrödinger equation solver (plus Fermi-Dirac weighting factors) for a distribution is that the actual distribution in the active region is a very non-equilibrium distribution which we must find as part of the modeling problem. But, the approaches discussed below are developed from the Schrödinger equation; we are solving a more general function, which incorporates the solution of this equation for an entire functional set. In addition, since the Schrödinger equation just defines a wave function, which is one part of the density (or a representation for an electron), it is quite difficult to incorporate dissipation through scattering mechanisms. Nevertheless, the starting point for all of our approaches lies in a mixed state wave function $\Psi(\mathbf{x},t)$, which is taken to be a field operator describing the degree of excitation of the various states of the system (this is one method of conveniently describing the mixed state of the system). Depending upon the Hamiltonian, this wave function can be a one-electron wave function or a many-body wave function.

A. The Density Matrix

The density matrix is formed from the composite of two such wave functions described above. It may be written as

$$\rho(\mathbf{x},\mathbf{x}',t) = \Psi(\mathbf{x},t)\Psi^+(\mathbf{x}',t), \quad (2)$$

where the "+" symbol on the second wave function indicates the Hermitian adjoint function. This is an *equal time* function and describes the correlation between events at positions \mathbf{x} and \mathbf{x}' . Obviously, $\langle \rho(\mathbf{x},\mathbf{x},t) \rangle = \langle \Psi(\mathbf{x},t)\Psi^+(\mathbf{x},t) \rangle = n(\mathbf{x},t)$ defines the local density of particles. Here, we have taken an expectation of the density *operator*, since the definition in (2) is obviously that of an operator. The equation of motion arises from the Liouville equation. It may be written as (in the absence of dissipative processes)

$$i\hbar \frac{\partial \rho}{\partial t} = -\frac{\hbar^2}{m} \frac{\partial^2 \rho}{\partial \mathbf{R} \partial \mathbf{s}} + 2 \left[\sinh\left(\frac{1}{2} \mathbf{s} \cdot \nabla\right) V \right] \rho, \quad (3)$$

where the last term, in the square brackets, is a short-hand notation for

$$\left[\sinh\left(\frac{1}{2} \mathbf{s} \cdot \nabla\right) V \right] = \frac{1}{2} \left[V(\mathbf{R} + \frac{\mathbf{s}}{2}) - V(\mathbf{R} - \frac{\mathbf{s}}{2}) \right], \quad (4)$$

and we have introduced the coordinate transformations

$$\mathbf{R} = \frac{1}{2}(\mathbf{x} + \mathbf{x}') \quad , \quad \mathbf{s} = \mathbf{x} - \mathbf{x}' \quad . \quad (5)$$

The density matrix has been used directly to study a number of devices. In Fig. 2(a), we show the density matrix for a double-barrier resonant tunneling diode (DBRTD) in equilibrium. For comparison, we also show the Wigner distribution function (described below) for this structure in Fig. 2(b). Both calculations are for barriers 0.3 eV high, 5 nm wide, and separated by 5 nm. Both are within lightly doped regions adjacent to the barriers, 5 nm wide for the Wigner function and 7.5 nm wide for the density matrix. In both calculations, the nominal density was 10^{18} cm^{-3} , and Fermi statistics were applied at the boundaries. For the density matrix in Fig. 2(a), this is represented by a damped oscillation in the nonlocal coordinate, whose period decreases as the density increases. Density is obtained from the diagonal component, which for the DBRTD, shows a small buildup of charge within the quantum well. The peak of this charge is approximately $2 \times 10^{16} \text{ cm}^{-3}$.

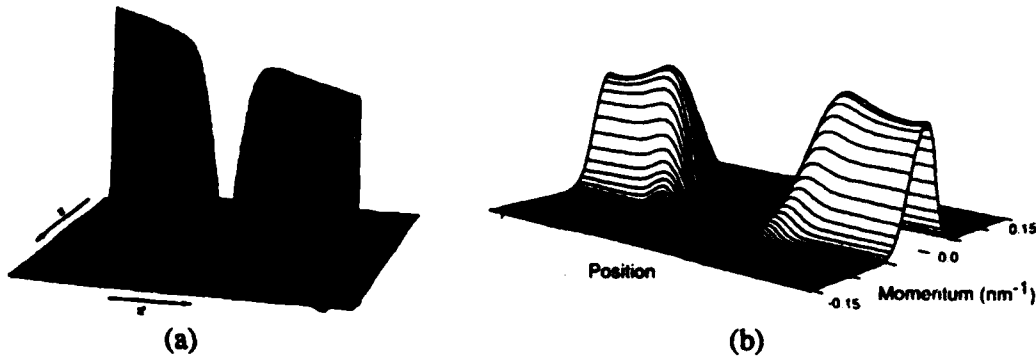


Figure 2. The double-barrier resonant tunneling diode in equilibrium: (a) the density matrix [10], (b) the Wigner distribution function [13].

For the density matrix under dynamic current flow conditions, dissipation is incorporated and serves to couple the real and imaginary parts of the density matrix. Current boundary conditions are represented by a *displaced distribution function*, similar to a displaced Fermi distribution. Dissipation is introduced as a phenomenological scattering potential whose diagonal components have the properties of a dynamic quasi-Fermi level [11]. At low values of bias, the scattering potential has the form $(\mathbf{x} - \mathbf{x}') [J/\tau \rho(\mathbf{x}, \mathbf{x})] \rho(\mathbf{x}, \mathbf{x}')$, which is similar to that discussed in [12]. Here J and τ represent current and scattering time, respectively. This form of scattering conserves the total number of particles. For simple barriers, the current-voltage characteristics display the expected exponential dependence on potential energy, with accumulation at the emitter side of the barrier and depletion on the collector side.

The computational procedures are described in detail in [12], and briefly may be described as re-expressing (3) as a coupled first-order system of equations, and seeking solutions along characteristic directions for the coupled equations. All of the calculations incorporate equally spaced grids, and a coupled Poisson solver. For simple barriers, the current-voltage characteristics display the expected exponential dependence on potential energy, with accumulation at the emitter side of the barrier and depletion on the collector side. This is shown in Fig. 3 for a 0.3 eV barrier, 15 nm thick, which is embedded in a 30 nm lightly doped region (GaAs). In Fig. 3(a), the real part of the density matrix, which is symmetric about the diagonal and shows charge accumulation on the emitter side. The imaginary part is shown in

Fig. 3(b), and is anti-symmetric, as its derivative along the nonlocal direction yields the current. We illustrate the computed current-voltage relationship in Fig. 3(c).

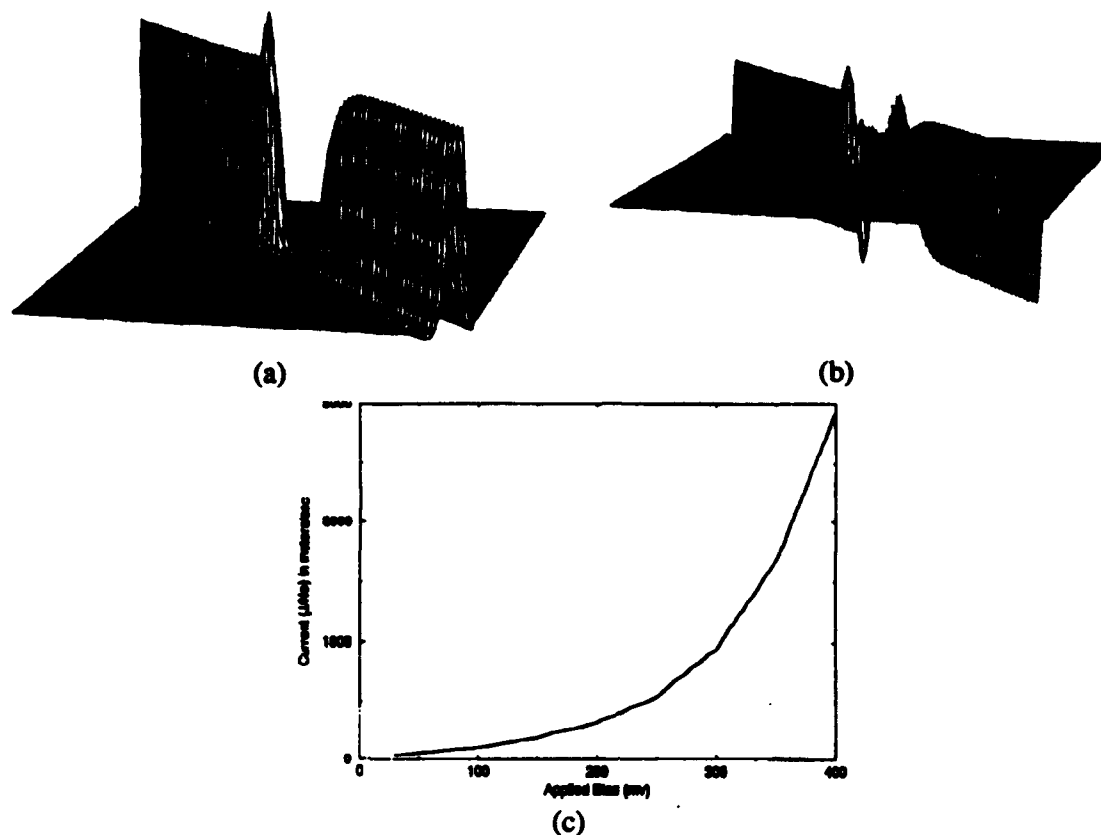


Figure 3. The DBRTD under bias, near the valley of the I-V curve. Parts (a and b) illustrate the real and imaginary parts of the density matrix, respectively, while the I-V curve itself is shown in (c).

B. The Wigner Distribution Function

The Wigner distribution becomes important when the physical problem is one that is better understood in terms of a phase-space distribution, and the carrier distribution function in an inhomogeneous device is one such problem. This phase-space distribution is not easily represented by the density matrix itself, but the Wigner distribution attempts to present an analogy between quantum and classical phase space for statistical mechanics. Since the statistical picture in phase space is well understood, indeed uses the Boltzmann equation for classical mechanics, transforming to a similar picture in quantum statistical mechanics allows the physical picture of a problem to be better understood. Unfortunately, position and momentum do not commute in quantum mechanics, and the two cannot be measured simultaneously to any great accuracy in phase space. This appears in the Wigner picture by regions of the phase space in which the distribution is negative in value. When the distribution is integrated over all space, the probability density in momentum space is recovered, and this quantity is positive definite. When the distribution is integrated over all momentum, the probability density in real space is recovered, and this quantity is also positive definite. In fact, if the Wigner distribution function is coarse-grain averaged over a region of phase space corresponding to a six-dimensional volume element whose size is set

by the uncertainty principle, the result is again a positive-definite "averaged" function. The Wigner distribution function is defined from the density matrix through the Fourier transform, often called a Weyl transform,

$$F_W(R, p, t) = \frac{1}{(2\pi\hbar)^3} \int d^3s \, e^{ip \cdot s / \hbar} \langle \rho(x + \frac{s}{2}, x - \frac{s}{2}) \rangle, \quad (6)$$

which in a sense is a Fourier transform on the variable that measures the distance from the diagonal in the density matrix $\rho(x, x')$. Since the Wigner distribution is a c-number, an expectation has been indicated in (6). This transformation accentuates the correlation that exists in the wave functions separated in position (if the correlation exists). The Wigner function builds in the correlations between different positions that are inherent in the off-diagonal elements of the density matrix. The Wigner function is evaluated at position R , but the density matrix terms that are used in the Fourier transform are those at the two positions $R \pm (s/2)$. The wave function may actually vanish at R , but the Wigner function will have a nonzero value in these areas in which the wave function vanishes, and the values in these such regions are measures of the correlation between the two endpoints on the vector s . The equation of motion for the Wigner distribution function is given by (again, in the absence of dissipation)

$$\frac{\partial F_W}{\partial t} - \frac{1}{m} p \cdot \nabla F_W = \frac{1}{\pi\hbar} \int d^3p' M(R, p') F_W(R, p + p', t), \quad (7)$$

where

$$M(R, p') = \int d^3s \, e^{ip' \cdot s / \hbar} [\sinh(\frac{1}{2} s \cdot \nabla) V]. \quad (8)$$

Equation (7) is quite similar to the streaming terms of the Boltzmann equation, especially if the lowest order term in the expansion of the potential is used. The Wigner distribution has also been used to model the DBRTD [13], and the results are shown in Fig. 4, again for the use of a relaxation time approximation for the dissipation, and for a bias near the valley of the I-V relation.

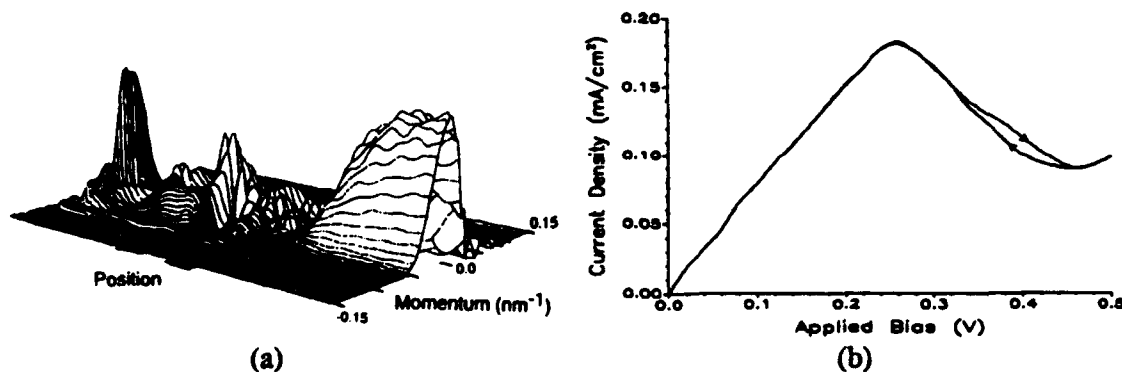


Figure 4. (a) Steady-state Wigner distribution at the valley of the I-V curve. (b) The I-V curve for a DBRTD.

The Wigner function in this simulation shows a depletion region in the cathode area, which arises from a contact potential drop and the tendency to form a bound state in this area. It is largely eliminated if a lightly-doped region is introduced adjacent to the barrier layers [13]. Such contact potential drops are typical of most open systems, whether classical or quantum, and are well-known in the Gunn effect device literature [14]. The depletion was also found

to go away with greater amounts of scattering in detailed studies of the role of scattering on the DBRTD by Frensky [15]. Generally, the cathode "barriers" will develop when there is a mismatch between the injection characteristics of the cathode reservoir and the dissipative nature of the active device region. Added dissipation or additional resistance in the active layer (through the lightly-doped regions) reduces the mismatch, and thereby reduces the depletion in the cathode-drop region.

C. Real-Time Green's Functions

Both the density matrix and the Wigner distribution function are *equal-time* functions, and are functions of only seven variables—either two vector positions and the time or the vector position, vector momentum, and the time, respectively. The energy does not enter into either description. This resulted from the definition that was used in (2), but there was no real requirement to have defined it in this manner. We could as easily have written the two wave functions at different times, and it is possible to define another function, which is a function of two times, through

$$G^<(x,t,x',t') = -i\langle\Psi^+(x',t')\Psi(x,t)\rangle. \quad (9)$$

The equal time version of (9) is obviously related to the density matrix itself. There are in fact a group of real-time Green's functions, arising from the different ways in which the wave functions can be combined and temporally ordered. Any simulation problem must solve for the independent members of this group (four in number) [16]. The particular Green's function in (9) is the "less than" function, and is closely related to the density in the other distributions above. Introducing the change of variables (5), and the equivalent for the average time T , and the difference time τ , we can introduce the energy \mathcal{E} ($=\hbar\omega$) through an additional Fourier transform, as

$$G^<(R,p,\omega,T) = i \int e^{ip\cdot s/\hbar - i\omega\tau} \langle\Psi(R + \frac{s}{2}, T + \frac{\tau}{2})\Psi^+(R - \frac{s}{2}, T - \frac{\tau}{2})\rangle d^3s d\tau. \quad (10)$$

It is clear that the $\tau = 0$ limit of (10) will lead to our Wigner distribution function, and

$$F_W(R,p,T) = \int d\omega G^<(R,p,\omega,T). \quad (11)$$

Much more information is contained within the Green's function formalism, since we can now investigate in detail the spectral density itself, which relates the energy to the momentum. However, very little has been done with these Green's functions in actual device modeling. However, they have been used to study high-field transport in homogeneous systems [16], and simplified versions have been used to study the DBRTD [17] for a non-self-consistent potential and weak scattering from phonons (but introduced without resorting to a relaxation time approximation). Nevertheless, the results are suggestive and indicate that quite detailed quantum device modeling can be carried out with the real-time Green's functions.

IV. Conclusions

Over the past few years, many groups have begun to explore quantum methodologies for modeling real semiconductor devices (at real temperatures). Most of the various approaches are closely related to each other, and offer different ways of approaching any given problem. While none of the techniques has become well developed, all of the ones discussed here have

been shown to lead to useful results and have given added insight into the problem. We are now passing the point at which we are trying to understand the methodology, and are in a position where we can now confidently use the methods to study device physics. Even so, many problems of understanding, particularly in the quantum statistical mechanics interpretations still remain, and will lead to many interesting lines of inquiry in the near future.

Acknowledgements

DKF would like to acknowledge support from the Office of Naval Research and the Army Research Office, while HLG would like to acknowledge support from the Office of Naval Research and the Air Force Office of Scientific Research.

References

- [1] J. Han, D. K. Ferry, and P. Newman, *IEEE Electron Dev. Lett.* **11**, 209 (1990).
- [2] A. Ishibashi, K. Funato, and Y. Mori, *Jpn. J. Appl. Phys.* **27**, L2382 (1988).
- [3] S. Y. Chou, D. R. Allee, R. F. Pease, and J. S. Harris, Jr., *Proc. IEEE* **79**, 1131 (1991).
- [4] T. Ando, F. Stern, and A. B. Fowler, *Rev. Mod. Phys.* **54**, 437 (1982).
- [5] A. M. Kriman, N. C. Kluksdahl, and D. K. Ferry, *Phys. Rev. B* **36**, 5953 (1987).
- [6] D. K. Ferry and J. R. Zhou, *Phys. Rev. B*, in press.
- [7] J.-R. Zhou and D. K. Ferry, *IEEE Trans. Electron Dev.* **39**, 473 (1992); **39**, 1793 (1992).
- [8] H. L. Grubin and J. P. Kreskovsky, *Sol.-State Electron.* **32**, 1071 (1989).
- [9] M. A. Ancona and G. J. Iafrate, *Phys. Rev. B* **39**, 9536 (1989).
- [10] G. J. Iafrate, H. L. Grubin, and D. K. Ferry, *J. Physique (Colloq. C-10)* **42**, 307 (1981).
- [11] T. R. Govindan and H. L. Grubin, to be published.
- [12] H. L. Grubin, T. R. Govindan, J. P. Kreskovsky, and M. A. Stroschio, *Sol.-State Electron.*, in press.
- [13] N. C. Kluksdahl, A. M. Kriman, D. K. Ferry, and C. Ringhofer, *Phys. Rev. B* **39**, 7720 (1989).
- [14] M. P. Shaw, H. L. Grubin, and P. R. Solomon, *The Gunn-Hilsum Effect* (Academic Press, New York, 1979).
- [15] W. R. Frensley, in *Computational Electronics*, Ed. by K. Hess, J. P. Leburton, and U. Ravaioli (Kluwer, Norwell, MA, 1991) 195.
- [16] For a review, see the chapters by G. Mahan and A.-P. Jauho in *Quantum Transport in Semiconductors*, Ed. by D. K. Ferry and C. Jacoboni (Plenum Press, New York, 1992).
- [17] R. Lake, G. Klimeck, M. McLennan, and S. Datta, in *Proc. Intern. Workshop on Computational Electronics* (Univ. Ill. Press, Urbana, 1992) 265; R. Lake and S. Datta, *Phys. Rev. B* **45**, 6670 (1992).

Boundary Conditions for Quantum Devices with Exposed Surfaces

Minhan Chen and Wolfgang Porod
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556, USA

David J. Kirkner
Department of Civil Engineering and Geological Sciences
University of Notre Dame
Notre Dame, IN 46556, USA

Abstract

The purpose of this work is to investigate the boundary conditions for the potential at exposed semiconductor surfaces in split-gate structure. A two dimensional numerical approach is presented for the coupling between the non-linear Poisson equation in the semiconductor (Finite Element Method) and Laplace's equation in the dielectric (Boundary Element Method). The utility of the coupling algorithm is demonstrated by simulating the potential distribution in a $n - \text{AlGaAs}/\text{GaAs}$ quantum wire structure with a semi-classical Thomas-Fermi charge model.

I. Introduction

Recent advances in nanostructure fabrication have made it possible to fabricate structures in which a two-dimensional layer of electrons is further confined into quantum wires or dots. Typically, such device structures are defined by metallic split gates. In order to understand the potential distribution in those structures, we solve Poisson's equation in the two-dimensional problem domain,

$$\epsilon \nabla^2 \phi = -\rho, \quad (1)$$

where ϕ is the electrostatic potential, ϵ is the dielectric constant, and ρ is the charge density [1]. Since this is a boundary value problem, one needs to know the values of the potentials and/or fluxes at the boundary. This is a crucial problem, especially at the exposed semiconductor surface.

In recent studies [1]-[7], the problem domain is typically identical with the semiconductor region, schematically shown in Fig. 1. The commonly used model for the boundary conditions on the exposed semiconductor surface is either a Dirichlet boundary condition [2],[3] or a Neumann boundary condition [4],[5]. Obviously, both of these models have their limitations, especially for very narrow split-gate structures used for quantum devices.

In this paper, we adopt an alternative viewpoint and develop an algorithm to implement boundary conditions at exposed semiconductor surfaces. We view as the natural problem domain the semiconductor *and* the dielectric, as shown in Fig. 2. Thus the artificial boundary conditions at the exposed surface are replaced by more physical matching

conditions at the interface between the semiconductor and the dielectric. Specifically, this algorithm, referred to as FBEM, couples a Finite Element solution of Poisson's equation in the semiconductor to a Boundary Element solution [8] of Laplace's equation in the dielectric with matching conditions.

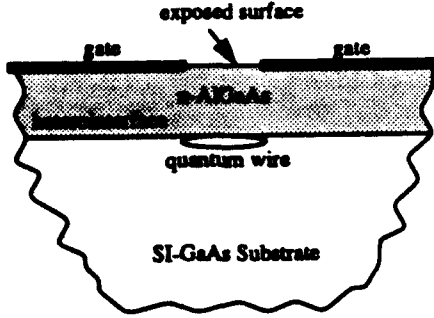


Figure 1: Typical problem domain for quantum devices defined by metallic gates. Shown are the multi-layer semiconductor regions and the exposed surface between the gates.

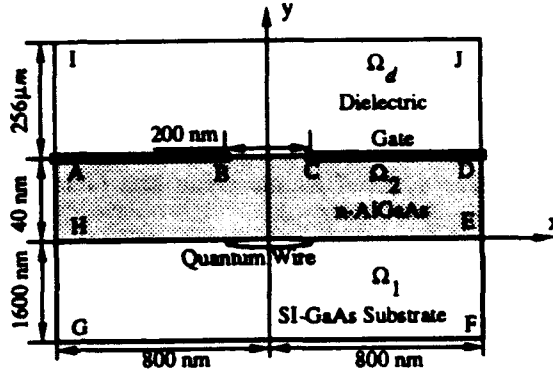


Figure 2: The problem domain consisting of the semiconductor regions, Ω_1 and Ω_2 , and the dielectric region, Ω_d . the semiconductor exposed surface B-C is treated as the interface between the semiconductor and the dielectric.

In Section II, we present the problem statement with the formulation of the boundary conditions on the exposed surface. In Section III, we discuss our numerical problem formulation. In Section IV, we present example results of the FBEM calculation for a $n - AlGaAs/GaAs$ quantum wire structure under bias conditions, and we compare them to those obtained with the usual Dirichlet and Neumann boundary conditions.

II. Problem Statement

We solve the two dimensional potential problem in the domain shown in Fig. 2, where regions Ω_1 with boundary $\partial\Omega_1$ and Ω_2 with boundary $\partial\Omega_2$ are semiconductor domains (non-linear Poisson equation) and region Ω_d with boundary $\partial\Omega_d$ is the dielectric domain (Laplace's equation). The simulation variable, u , is defined as the potential difference between the conduction band edge E_C and the Fermi energy E_F in units of the thermal energy kT , i.e., $u \equiv (E_C(\phi) - E_F)/kT = (E_C^b - e\phi - E_F)/kT$, where E_C^b is the conduction band edge in the bulk. The boundary C-B is the exposed semiconductor surface. Mathematically C-B is the interface between the region Ω_2 and the region Ω_d . Across this interface, the potential, u , is continuous and the jump in the normal electric flux density is equal to the interface charge density. The problem to be solved can be posed as follows:

Find:

$$u = u_1, u_2, u_d, \quad (2)$$

such that

$$\epsilon_l \nabla^2 u_l(x, y) = -f[u_l(x, y)], \quad (x, y) \in \Omega_l, \quad l = 1, 2, \quad (3)$$

$$\epsilon_d \nabla^2 u_d(x, y) = 0, \quad (x, y) \in \Omega_d, \quad (4)$$

and with interface matching conditions at the exposed surface, $\partial\Omega_{BC}$:

$$u_2 - u_d = 0, \quad \text{on } \partial\Omega_{CB}, \quad (5)$$

$$\epsilon_2 \frac{\partial u_2}{\partial \vec{n}} - \epsilon_d \frac{\partial u_d}{\partial \vec{n}} = \frac{e}{kT} Q_{int}, \quad \text{on } \partial\Omega_{CB}, \quad (6)$$

where $f(u_i) = f_i = e\rho(u_i)/kT$, is the charge density term in the domain Ω_i , and Q_{int} is the interface charge density on the interface $\partial\Omega_{CB}$. Generally Q_{int} and ρ may be a function of the potential u , that is, $Q_{int} = Q_{int}(u)$ and $\rho = \rho(u)$.

III. Numerical Formulation

The semiconductor domain Ω_s , $\Omega_s = \Omega_1 \cup \Omega_2$, where the non-linear Poisson equation (3) governs, is discretized in a way suitable for the application of the Finite Element Method (FEM). The resultant non-linear system of equations is:

$$\begin{aligned} K_{11}u_o^s + K_{12}u_{BC}^s &= p_f^s, \\ K_{12}^T u_o^s + K_{22}u_{BC}^s &= p_{BC}^s, \end{aligned} \quad (7)$$

where u_{BC}^s and p_{BC}^s contain the potentials and nodal forces, respectively, at the nodes on the interface $\partial\Omega_{BC}$ between the semiconductor and the dielectric, u_o^s and p_f^s contain the potential and nodal forces at all other nodes in the semiconductor domain, respectively, and K is the stiffness matrix.

The dielectric domain, Ω_d , is a homogeneous charge free region. The governing equation is Laplace's equation. Since the fundamental solution of Laplace's equation (4) is known, a boundary integral equation technique can be employed. The resultant system of equations can be expressed as:

$$\begin{aligned} S_{11}u_o^d + S_{12}u_{BC}^d &= p_o^d, \\ S_{21}u_o^d + S_{22}u_{BC}^d &= p_{BC}^d. \end{aligned} \quad (8)$$

where S^d is the equivalent stiffness matrix and p^d is the equivalent nodal force vector.

Writing the matching conditions at the exposed surface, (5) and (6), in discretized form,

$$u_{BC}^s = u_{BC}^d = u_{BC}, \quad (9)$$

$$p_{BC}^s + p_{BC}^d = q. \quad (10)$$

A new global system of equations is formed by coupling the dielectric, equations (8), with the semiconductor, equations (7), and enforcing the matching conditions (9) and (10). vspace-0.1in

$$\begin{pmatrix} S_{11} & S_{12} & 0 & 0 \\ S_{21} & S_{22} & 0 & I \\ 0 & K_{12} & K_{11} & 0 \\ 0 & K_{22} & K_{12}^T & -I \end{pmatrix} \begin{pmatrix} u_o^d \\ u_{BC}^d \\ u_o^s \\ p_{BC}^s \end{pmatrix} = \begin{pmatrix} p_o^d \\ q \\ p_f^s \\ 0 \end{pmatrix}. \quad (11)$$

Solution of this set yields the potential distribution in the semiconductor domain, including the interface $\partial\Omega_{BC}$, and the nodal flux on $\partial\Omega_{BC}$.

IV. Example

The example structure with its dimensions is shown in Fig. 2. The quantum wire is realized at the $Al_{0.3}Ga_{0.7}As/GaAs$ heterojunction and is defined by the metal gates on the top surface. The gate electrodes have an applied voltage V_g . Between the gates is the exposed semiconductor surface where the interface charge density Q_{int} is assumed to be fixed for the example although this is not a limitation of the algorithm. The n -type doping density is assumed to be $10^{18}cm^{-3}$ in the $Al_{0.3}Ga_{0.7}As$ layer and $10^{15}cm^{-3}$ in the $GaAs$ substrate. A semi-classical Thomas-Fermi charge model [2, 7] is assumed in the semiconductor domain.

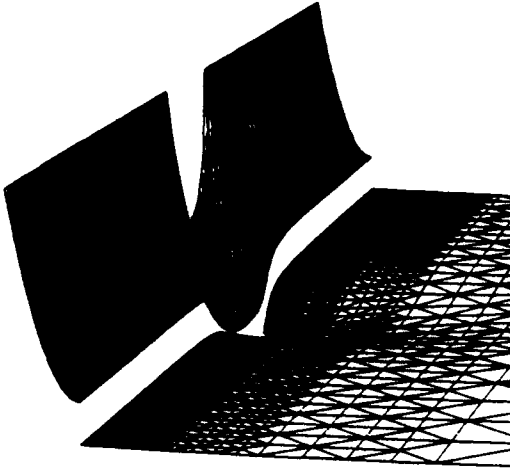


Figure 3: Sample result of the potential landscape in the semiconductor region. Shown is the conduction band obtained by the FBEM algorithm for -1.0V gate bias. The heterointerface is clearly visible at the discontinuity of the potential.

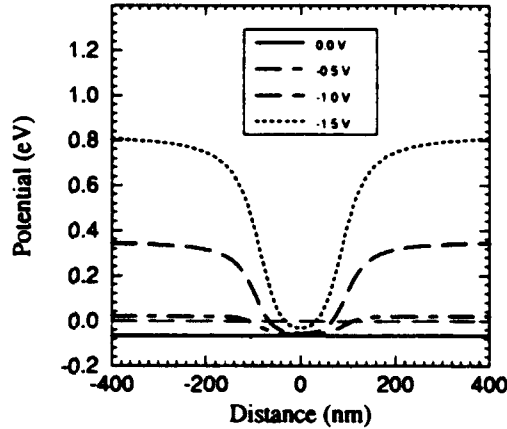


Figure 4: Result of the FBEM algorithm showing the conduction band profiles parallel to the heterointerface (on the GaAs side) for different bias conditions.

As shown in the center of Fig. 3, a potential well is formed underneath the exposed surface for a negative gate bias of -1.0 V. In Fig. 4, the conduction band is plotted parallel to the $AlGaAs/GaAs$ heterointerface (on the GaAs side) for different bias conditions. An electron gas is formed in those regions where the conduction band dips below the semi-classical electron quasi Fermi level, which is chosen as the zero of the energy axis and indicated by the dashed line. Figure 5 shows a comparison of the results obtained with our FBEM algorithm and those utilizing the conventional Dirichlet or Neumann boundary conditions. Shown is the conduction band profile parallel to the $AlGaAs/GaAs$ heterointerface (on the GaAs side) in Fig. 5(a), and the different methods produce significantly different results. Specifically, the expected width of the electron channel differs by a factor of two for the Dirichlet and FBEM boundary conditions. The Dirichlet boundary conditions also predict a significantly higher electron density than the FBEM algorithm. Figure 5(b) shows the same comparison at the semiconductor surface. Both the Dirichlet and Neumann boundary conditions exhibit an unrealistic discontinuous behavior of the potential, as opposed to the physically more appealing smooth result

of the FBEM algorithm.

In summary, we have presented a study of the boundary conditions at exposed semiconductor surfaces and developed an algorithm, termed FBEM, to solve this type of potential problem in the semiconductor and the dielectric. The major advantage of our algorithm is to model the exposed semiconductor surface by imposing the more physical interface matching conditions without making artificial assumption about either the potential or the electric field at the exposed surface. In ongoing work, we study the implementation of different physical models for the interface charge Q_{int} at the exposed surface. These results will be presented in the future.

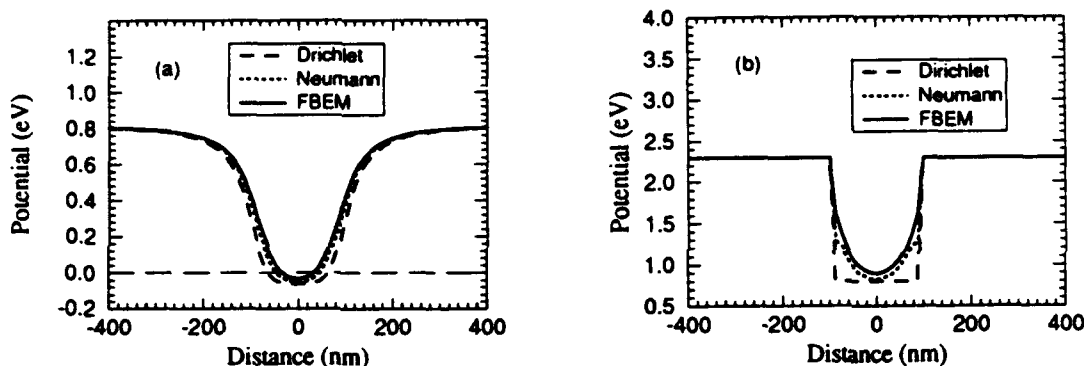


Figure 5: Comparison of the conduction band profiles under -1.5 V gate bias for the three types of boundary conditions on the exposed semiconductor surface; (a) parallel to the heterointerface (on the GaAs side), and (b) parallel to the semiconductor surface.

Acknowledgements

The authors wish to thank Dr. Gary H. Bernstein, Mr. Henry K. Harbury, Dr. Craig S. Lent, Mr. Mark Mueller, and Mrs. Helen Zhang for numerous helpful discussions and suggestions. This work was supported in part by AFOSR, ARPA and ONR.

References

- [1] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*, (Springer-Verlag, 1984).
- [2] J. H. Luscombe, *Nanotechnology* **4**, 1 (1993).
- [3] A. Bouchard and J. Luscombe, *Nanostructures and Mesoscopic Systems*, ed. by W. P. Kirk and M. A. Reed, pp. 393 (Academic, 1992).
- [4] S. E. Laux, NASECODE V: Proc. 5th Intl. Conf. on "Numerical Analysis of Semiconductor Devices and Integrated Circuits", pp. 270 (1987).
- [5] A. Kumar, S. E. Laux and F. Stern, *Phys. Rev. B* **42**, 5166 (1990).
- [6] U. Ravaioli, T. Kerkhoven, M. Raschke and A. T. Galick, *Superlatt. Microstruct.* **11**, 343 (1992).
- [7] W. Porod, H. K. Harbury and S. M. Goodnick, *Appl. Phys. Lett.* **61**, 1823 (1992); and H. K. Harbury, W. Porod and S. M. Goodnick, *J. Appl. Phys.* **73**, 1509 (1993).
- [8] C. A. Brebbia, J. C. F. Telles, and L. C. Wrobel, *Boundary Element Techniques*, (Springer-Verlag, 1984).

Queuing-theoretic simulation of single-electronic metal-semiconductor devices and systems

Sharif Babiker, John R. Barker and Asen Asenov

Nanoelectronics Research Centre
Department of Electronics and Electrical Engineering
University of Glasgow, Glasgow G12 8QQ, UK

Abstract

Traffic theoretic or queuing methods are proposed as a natural framework for modelling the correlated tunnelling and transport of single electron solitons in coupled tunnel capacitor structures below the Coulomb blockade threshold. Stationary state and time-dependent modelling is developed and validated against Monte Carlo simulation. Exact results are obtained for the double tunnel junction or quantum dot. The effects of discrete energy states are evaluated and the extension to multiple junctions illustrated.

I. Introduction

The controlled transport/correlated-tunnelling of single-electronic excitations in coupled tunnelling capacitor structures is now experimentally established in metal-insulator, metal-semiconductor systems and in capacitatively-coupled quantum point contacts in semiconductor 2DEG structures[1-4]. These structures rely on the existence of ultra-small capacitive structures such that the effective charging energies $e^2/2C$ exceed the thermal energy $k_B T$ [5]. By exploiting state-of-the-art nanofabrication it is possible to construct 20nm scale coupled capacitors (metal on semiconductor coupled Schottky dots)[6] which point the way to a future high temperature, high density nanoelectronic systems technology (Fig.1).

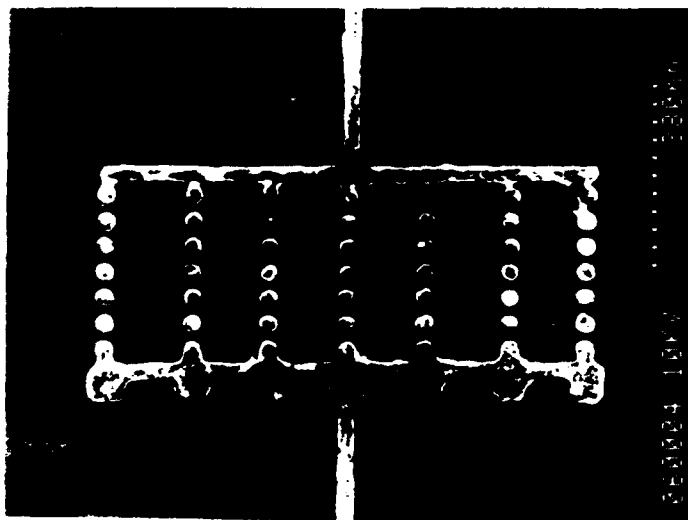


Figure 1. 40 nm diameter hemispherical Aluminium Schottky dots on p-Silicon with 15 nm spacing (Weaver[6]).

Single-electronic, ultra-small capacitor systems provide a new regime of study for semiconductor device modelling wherein the effects of charging become important and the Coulomb interaction provides subtle correlation effects.

II. Monte Carlo method

Simulations of arrays of single-electronic tunnel junctions (see figure 2) and gated arrays (figure 3) have been carried out to date by Monte Carlo methods which have proved to be prohibitive computationally for the extended systems of technological interest.

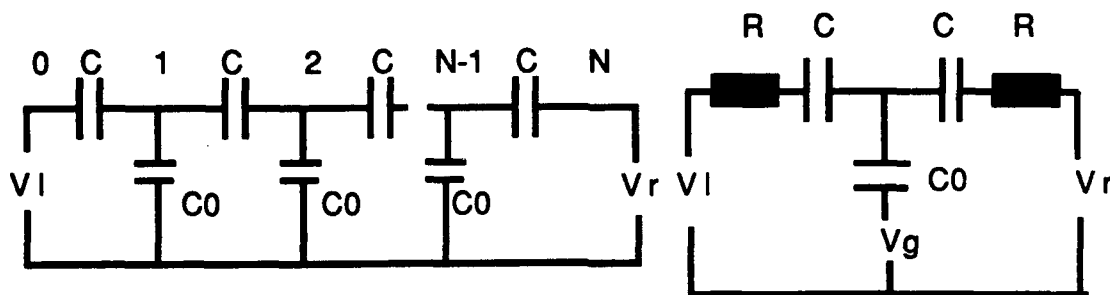


Figure 2. A multi-junction tunnelling array. The capacitances C are tunnel junctions. C_0 is taken to be non-tunnelling.

Figure 3. Gated 2-junction array

In the Monte Carlo approach[7,8] the state of the tunnel junction array is described by the number of solitons (electron plus polarisation field) at each node. The total charge on each junction/tunnel junction/ground capacitance is the sum of the charges due to the voltage sources and the charges induced by the soliton structure. These charges are linearly related to the voltages and number of solitons at each node. The evolution of the array is then determined by the stochastic process of single electron tunnelling which alters the soliton occupancy vector \mathbf{n} .

III. Queuing theory: stationary case

In the present paper we introduce a new approach to modelling single-electronic systems that captures the Poisson stochastic nature of tunnel events and provides a fast, physically transparent and efficient method of calculating the steady-state characteristics of multi-junction configurations. This new method involves a re-formulation of the transport equations in terms of *queuing theory* (traffic theory[9]) and centres on determining the distribution P_i of quasi-electrostatic soliton excitations that are formed during the transport/tunnelling process. For an N junction array the state of the system is described by $S_i = (k_1, k_2, \dots, k_N)$ where k_i are the number of excess electrons(solitons) at node k_i (figure 4 shows states for a 2-junction system).

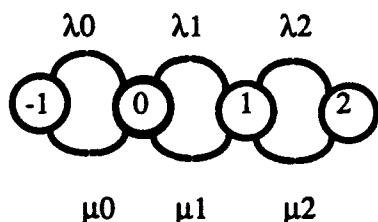


Figure 4. Soliton states in a 2-junction.

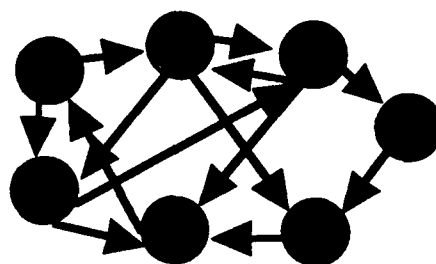


Figure 5. Transitions of soliton states

An electron may tunnel from a node to a neighbouring node provided a tunnelling connection exists (connection matrix $X_{ij} = 1$). Electrons form a queue at each arrival node where they wait for "service" (unlike conventional traffic theory the queues are always full and the "service rate"

which is equal to μ_i the total departure rate from each node i varies as a function of the state of the system). The current passing between two neighbouring nodes (l,k) which are connected by a tunnel junction with capacitance C_{lk} is calculated from the P_i by $I_{lk} = e \cdot \sum_i P_i (\Gamma_i(k,l) - \Gamma_i(l,k))$ where P_i is the probability of finding the system in state S_i , and $\Gamma_i(k,l) =$ tunnelling rate from node k to neighbouring node l when the system is at state S_i . The average voltage across the capacitor C_{kl} can also be calculated as $v_{kl} = \sum_i P_i \cdot q_{kl}(i) / C_{kl}$. To determine the distribution of the soliton states P_i , consider a single tunnel event from node k to node l at state S_i to form state S_j . Define the departure rate $\mu_{ij} = \Gamma_{ni}(k,l)$ for $X_{kl} = 1$, from state S_i to S_j for the system already occupying S_i . The Poisson nature of the tunnelling events allows us to find the total departure rate from state S_i as: $\mu_i = \sum_j \mu_{ij}$ and the transition probability from state S_i to S_j is thus, $r_{ij} = \mu_{ij} / \mu_i$ defining the routing matrix R . The average residence time (per visit) at state S_i is $1/\mu_i$. The input traffic to node i is given by $\lambda_i = \sum \lambda_l r_{li}$, or in matrix form, the traffic equations are: $\lambda \cdot (I - R) = 0$ where λ is the row vector $(\lambda_1, \lambda_2, \dots, \lambda_n)$. Unfortunately, the matrix $(I - R)$ has a zero determinant and therefore, the traffic equations, have an infinite number of solutions. To proceed let us define the occupancy vector $\underline{m} = (m_1, m_2, \dots, m_n)$ where $m_i = 1$ if the system is found at state S_i and $m_i = 0$ otherwise. It is clear that the elements of \underline{m} always satisfy $\sum_i m_i = 1$. The model described is similar to a closed network of servers in which there is exactly one job travelling (trapped) between the service centres. Let λ^* be some non zero solution of the traffic equations. The probability distribution of \underline{m} is given as [12]: $p(\underline{m}) = \alpha_1(m_1) \cdot \alpha_2(m_2) \cdot \dots \cdot \alpha_n(m_n) / G$, where $\alpha_i(0) = 1$ & $\alpha_i(1) = \lambda_i^* / \mu_i$. G is a normalisation constant evaluated as: $G = \sum \prod_i \alpha_i(m_i)$, where the summation is realised over all possible vectors \underline{m} . G reduces to $G = \sum_i \lambda_i^* / \mu_i$. Finally, the distribution of soliton structures is found as,

$$P_i = G^{-1} \cdot \lambda_i^* / \mu_i \quad \text{which is the key result of this paper.}$$

IV. Exactly solvable example: the double junction

Figure 6 shows an application to the 2-junction (or single quantum dot) compared with results obtained by the Monte Carlo method (750 events). Above a threshold voltage V_{th} electrons can tunnel into and away from the dot, one electron at a time. For N excess electrons the total electrostatic energy due to this charge is $(Ne)^2 / 2C_T$ where C_T is the total capacitance seen by the charge, $C_T = C_1 + C_2 + C_0$ and C_1, C_2 are the capacitances between the dot and the metallic electrodes and C_0 is the capacitance to the ground electrode. In this case, on leaving state S_i the system can only make a transition to states S_{i+1} or S_{i-1} (see figure 4). This is a Markovian birth-death process which can be solved exactly using the transition rates $\lambda_n (S_{n-1} \rightarrow S_n)$ and $\mu_n (S_n \rightarrow S_{n-1})$ given in terms of the tunnelling rates by: $\lambda_n = \Gamma_{n-1}(l,c) + \Gamma_{n-1}(r,c)$; $\mu_n = \Gamma_n(c,l) + \Gamma_n(c,r)$. The probability of finding the system at state S_n is found as:

$$P_n = Z^{-1} \cdot \prod_{i=k_1+1}^n (\lambda_i / \mu_i) \quad (n > k_1) \quad \text{or} \quad Z^{-1} \quad (n = k_1)$$

$$Z = 1 + \sum_{i=k_1}^{k_2} \prod_{n=k_1+1}^i (\lambda_n / \mu_n)$$

and where k_1 and k_2 are the minimum and maximum possible number of excess electrons that can be accommodated on the dot. The states contributing to the process can be discovered according to the condition: if $\lambda_i > 0$ and $\mu_i > 0$ then both S_i and S_{i-1} are legal states.

The oscillations in the dc conductance with gate voltage at low bias voltage correspond to transport through the same number of soliton states. In Fig 6(a) transport in the different segments corresponds to states (0,1), (1,2), (2,3) and so on. At higher bias more states are generated leading to splitting of the segments. In Fig 6(b) the sequences are: (0,1), (0,1,2), ((1,2) and (1,2,3). The maximum of the conductance peaks increases with applied bias whereas the minima decrease in amplitude. For high bias $V \gg V_{th}$, the conductance approaches the constant value $G \rightarrow G_t/2$. The speed-up over Monte Carlo simulation (750 departures) is about 300 x in CPU time.

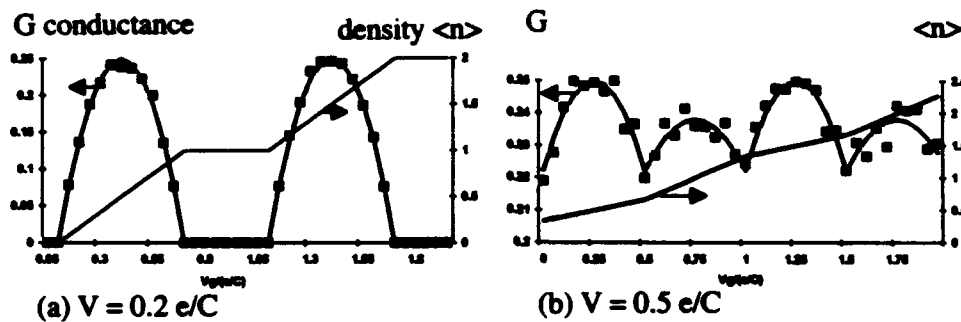


Figure 6. Conductances G/G_0 and soliton density $\langle n \rangle$ as a function of voltage V_g for 2 tunnel junctions at bias voltages $V=0.2, 0.5$ e/C. Squares: Monte Carlo .lines: queuing theory .

V. Time evolution

Suppose the system is found in state $S = S_k$ at $t = 0$ with initial densities $P_i(0) = \delta_{ik}$. At a later time t the system will be in a mixture of all possible legal states with densities $P_i(t)$ given by the rate equation: $\partial P_i / \partial t = \sum_j \{ P_j \mu_{ji} \} - P_i \mu_i$. As an application, consider a single quantum dot described by the 2 soliton state $\{S\} = \{n, n+1\}$, with birth and death coefficients λ and μ respectively. If the system is known to be in state S_n at $t=0$, the rate equation may be solved to give: $P_n(t) = \{ \mu + \lambda \exp(-(\lambda + \mu)t) \} / (\lambda + \mu)$; $P_{n+1}(t) = \lambda \{ 1 - \exp(-(\lambda + \mu)t) \} / (\lambda + \mu)$ which in the limit $t \rightarrow \infty$ recovers our earlier result $P_n(\infty) = \mu / (\lambda + \mu)$; $P_{n+1}(\infty) = \lambda / (\lambda + \mu)$.

VI. General application to multiple junction systems and quantum dots

Very complex behaviour is possible with soliton propagation in multiple junctions. Figure 7 shows an example from the traffic theory where we have enhanced the magnitude of the conductance oscillations in a 3-junction array by choosing one element to have a large RC value. Here $V_r = 0$ and $V_l = -V$. In region (a) a maximum of one excess electron can stay in the system; the possible states are (0,0), (0,1), (1,0) and the system stays mainly in state (0,1). In region (b) the dominany state is (1,1) and up to 2 electrons excess occur. In region (c) 11 states contribute to the conduction with between 0 and 3 excess electrons. Region (d) has 12 states with between -1 and 3 excess charges. The conductance evidently peaks each time the system can accommodate one more electron.

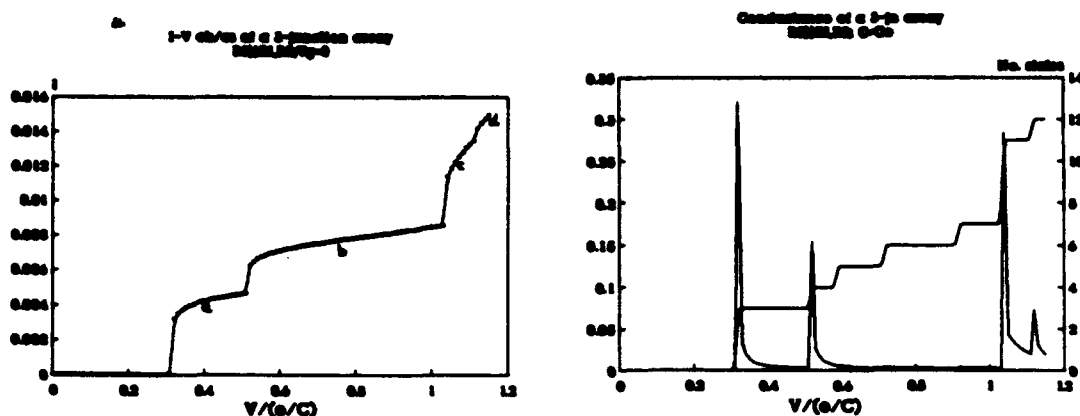


Figure 7. I-V characteristics, conductance and number of soliton states for a 3-junction having $R_3 = 50 R_1$, $R_1=R_2$; $C_1=C_2=C_3=C$; $C=C_0$.

The methodology has been extended to (a) quantum dot systems where the isolated charging islands have discrete energy levels[13]; (b) macroscopic quantum tunnelling processes [10] The theory is being applied to the simulation of ultra-small coupled Schottky dot structures[6] for which the capacitance matrix requires a full 3D Poisson solver.

References

- [1] T A Fulton and G J Dolan, Phys Rev Letters, **89** 109 (1987).
- [2] L S Kuzmin, P Delsing, T Claeson, K K Likhaev, Phys Rev Letters **60** 309 (1989), **62** 2539 (1989)
- [3] L J Geerligs, V F Anderagg, P A M Holweg, J E Mooij, H Pothier, D Esteves, C Urbina, M H Devoret, Phys Rev. Letters **64** 2691 (1990).
- [4] U Meir, M A Kastner and S J Wind Phys RevB **40** 5871 (1989).
- [5] D V Averin and K K Likhaev, J. Low Temp Phys **62** 345 (1986).
- [6] J R Barker, J M R Weaver, S Babiker and S Roy, *Theory, modelling and construction of single-electronic systems*, Proceedings Second International Symposium on New Phenomena in Mesoscopic Structures, Hawaii, (1992).
- [7] N S Bakhvalov, G S Kazach, K K Likhaev, S I Serdyukova, Sov. Phys. JETP **68** 581 (1989).
- [8] J R Barker, S Roy and S Babiker, *Trajectory representations, fluctuations and stability of granular electronic devices*, Chapter 22, pp213-232, Science and Technology of Mesoscopic Structures, edited by S. Namba, C. Hamaguchi, T. Ando. Springer-Verlag: London, Tokyo, New York. (1992)
- [9] L. Kleinrock, "Queuing Systems", New-York, London, (1974, 1976).
- [10] D. Averin & A. Odintsov, Phys. Lett.(A), **140**, 251 (1989).
- [11] M Schwartz, "Computer communication Network Analysis", Prentice-Hall Inc (1977).
- [12] J Jackson, "Networking of waiting lines", Operations Research, **518**, (1959).
- [13] C Beenakker, Phys. Rev B **44**, 1646 (1991).

SIMULATION OF A Si/SiGe MODULATION-DOPED FET USING QUANTUM HYDRODYNAMIC EQUATIONS*

J.-R. Zhou, T. Yamada, H. Miyata[†], and D. K. Ferry

Center for Solid State Electronics Research
Arizona State University, Tempe, AZ 85287-6206, U.S.A.

Abstract

We present here the simulation of a Si/SiGe modulation-doped HEMT. The electron transport properties were obtained from Monte Carlo simulation. We use a set of quantum hydrodynamic equations for the device simulation. The calculated transconductance is about 300 mS/mm at 300 K. Velocity overshoot in the strained Si channel is observed. The inclusion of the quantum correction increases the total current by as much as 15 per cent.

I. Introduction

Modulation doped $\text{Si}_{1-x}\text{Ge}_x/\text{Si}/\text{Si}_{1-x}\text{Ge}_x$ offers a Si-version of the HEMT. With achievable high mobility in the strained Si layer (2000 ~ 3000 $\text{cm}^2/\text{V}\cdot\text{s}$ at 300 K [1-2], 11,000 $\text{cm}^2/\text{V}\cdot\text{s}$ at 77 K [1], and 175,000 $\text{cm}^2/\text{V}\cdot\text{s}$ at 1.5 K [3-4]), the prospects of a high performance FET are good. Experimental devices have achieved transconductance as high as 330 mS/mm at 300 K and 600 mS/mm at 77 K in a 0.25 μm gate device [1]. In this paper, we present a numerical simulation of the modulation-doped SiGe device.

In preparation for device simulation, we have obtained electron transport properties, such as the velocity-field and energy-field characteristics for strained Si and $\text{Si}_{1-x}\text{Ge}_x$ material, by Monte Carlo simulation [2]. The device modeled here has a gate length of 0.18 μm and we use a set of quantum hydrodynamic equations which utilize these Monte Carlo results. The calculated transconductance is about 300 mS/mm at 300 K. Velocity overshoot in the strained Si channel is observed. The inclusion of the quantum correction increases the total current by as much as 15 per cent.

II. Modeling of Si/SiGe modulation-doped FET

A set of hydrodynamic equations, which is described below, is used in the simulation. The equations, which are essentially the same as those used in [5-7], describe the particle conservation, momentum conservation, and energy conservation. Written in the temperature representation, we have:

$$\frac{\partial n}{\partial t} + \nabla \cdot (n\mathbf{v}) = 0, \quad (1)$$

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} = -\frac{q\mathbf{E}}{m^*} - \frac{1}{nm^*} \nabla (nk_B T_q) - \frac{\mathbf{v}}{\tau_m}, \quad (2)$$

$$\frac{\partial T}{\partial t} + \frac{1}{3\gamma} \mathbf{v} \cdot \nabla (T_q) = -\frac{2}{3\gamma} \nabla \cdot (\mathbf{v} T_q) + \frac{m^* v^2}{3\gamma k_B} \left(\frac{2}{\tau_m} - \frac{1}{\tau_w} \right) - \frac{T - T_0}{\tau_w}, \quad (3)$$

* Work supported by the Army Research Office.

[†] On leave from Fujitsu Basic Processes Division, Kawasaki, Japan.

where n is the average electron density, v is the average electron velocity, T is the effective electron temperature, m^* is the effective electron mass, E is the electric field, τ_m is the momentum relaxation time, τ_w is the energy relaxation time, T_0 is the lattice temperature, and T_q is given by

$$T_q = \gamma T + \frac{2}{3k_B} U_q, \quad (4)$$

with

$$U_q = -\frac{\pi^2}{8m^*} \nabla^2 \ln(n), \quad (5)$$

and γ is the degeneracy factor [7]

$$\gamma = \gamma(\mu_f/k_B T) = \frac{F_{3/2}(\mu_f/k_B T)}{F_{1/2}(\mu_f/k_B T)}, \quad (6)$$

where F_j is the Fermi-Dirac integral, and μ_f is the Fermi energy measured from the conduction band edge. The factor γ is introduced as a correction to the total average electron kinetic energy (assuming a Fermi-Dirac distribution function):

$$w = \frac{1}{2} m^* v^2 + \frac{3}{2} \gamma k_B T + U_q. \quad (7)$$

Our Monte Carlo simulation was carried out to obtain the transport properties of the SiGe materials. The computed velocity-field and energy-field relations are plotted in Figs. 1 and 2 for strained Si grown on relaxed $\text{Si}_{0.7}\text{Ge}_{0.3}$, with modulation-doped concentrations of $1.5 \times 10^{18} \text{ cm}^{-3}$ and $1.0 \times 10^{14} \text{ cm}^{-3}$ at 300 K. The velocity curves in Fig. 1 show that, due to the higher mobility of the electrons in the strained Si, electrons have higher velocity than that in SiGe material. The low-field mobility is found to be about $3000 \text{ cm}^2/\text{V}\cdot\text{s}$, and a slightly larger saturation velocity is observed. The average electron energy (Fig. 2) for strained Si rises faster at low field because of the light transverse mass and reduced intervalley scattering, but is less at high field due to impact ionization.

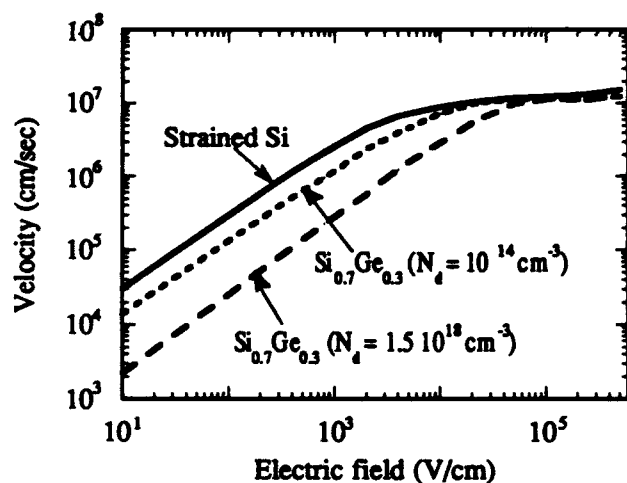


Figure 1. Velocity as a function of field for strained Si on relaxed SiGe, found from Monte Carlo simulation.

The relaxation times τ_m and τ_w , which are functions of energy, are determined by fitting the homogeneous hydrodynamic equations to the velocity-field and energy-field relations in Fig. 1 and Fig. 2.

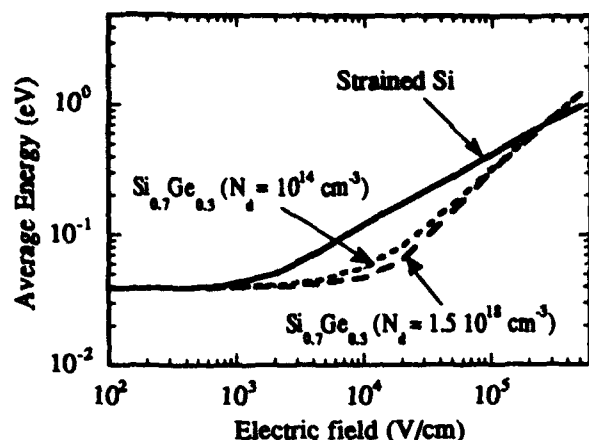


Fig. 2 Average energy as a function of field for strained Si on relaxed SiGe, found from Monte Carlo simulation.

The numerical simulation has been applied to a $0.18\ \mu\text{m}$ gate, quantum-well device with a modulation-doped structure of $\text{Si}_{0.7}\text{Ge}_{0.3}/\text{Si}/\text{Si}_{0.7}\text{Ge}_{0.3}$. The device structure is shown in Fig. 3. The doping of the top $\text{Si}_{0.7}\text{Ge}_{0.3}$ layer is taken to be $3.5 \times 10^{18}\ \text{cm}^{-3}$, and a doping of $1.0 \times 10^{14}\ \text{cm}^{-3}$ is used in the substrate $\text{Si}_{0.7}\text{Ge}_{0.3}$. The lattice temperature in the simulation is taken to be 300 K. The typical simulation domain is $1.0\ \mu\text{m} \times 0.095\ \mu\text{m}$. The thickness of the top SiGe layer is 19 nm, and the strained Si channel is 18 nm. The graded interface transition at the SiGe/Si junction results in an effective 3 nm spacer layer. For simplicity, we only use a three-layer structure. The modulation-doped structure in [1] is more complicated. However, the active structures are similar.

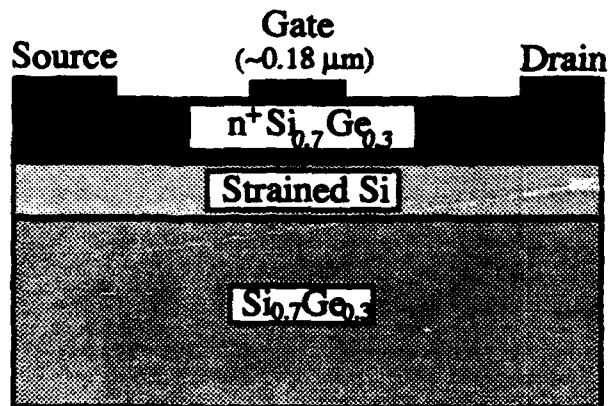


Fig. 3 The device structure.

III. The simulation results

The computed I-V characteristics for the $0.18\ \mu\text{m}$ gate device are shown in Fig. 4, for gate biases of 0.7, 0.5, 0.2, and 0 volts, respectively. The small thickness of the top SiGe layer provides a normally-off device, since a Schottky barrier height of 0.9 V leads to an estimated depletion width of 18.4 nm. The peak transconductance is about 300 mS/mm, and good saturation with a drain conductance of 4.6 mS/mm at the gate voltage of 0.5 V is obtained. Approximately the

same current level and transconductance was found in a $0.25\text{ }\mu\text{m}$ device. The simulation results are comparable to the experimented results in [1]. The relatively larger current level ($0.3\text{ mA}/\mu\text{m}$) and transconductance (330 mS/mm) found in the experiment is thought to be due to a higher sheet density ($2.5 \times 10^{12}\text{ cm}^{-2}$ [1] compared to $1 \times 10^{12}\text{ cm}^{-2}$ in this simulation) in the quantum well for their particular modulation-doped structure.

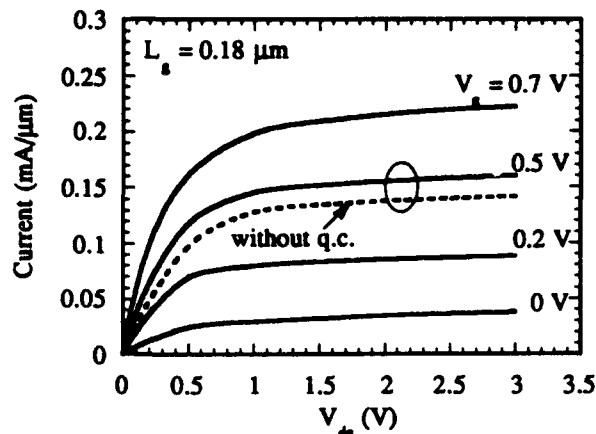


Fig. 4 I-V characteristics of a $0.18\text{ }\mu\text{m}$ SiGe device.

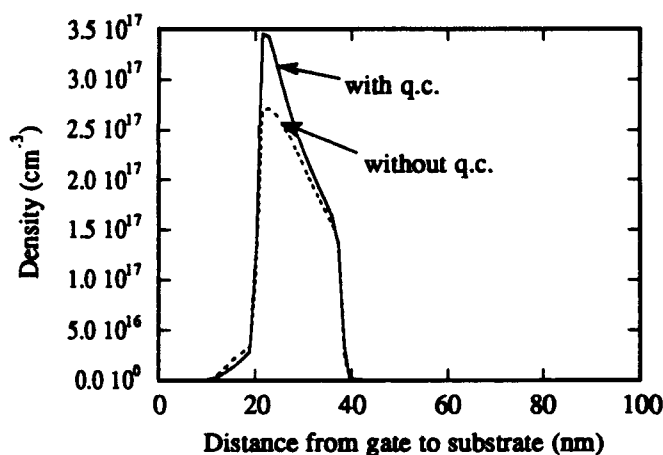


Fig. 5 Electron density across the channel.

Our simulation shows that, without including quantum corrections, the current would be 15 per cent smaller for the $0.18\text{ }\mu\text{m}$ gate device at a gate voltage of 0.5 V , as shown in Fig. 4. In other words, the inclusion of the quantum potential increases the total current by as much as 15 per cent in the simulation. This large modification of the current was not expected in the device with such a gate length. However, by inspecting the density distribution along the channel, one can find that a rapid density change occurs at the gate end close to the drain contact within a region much shorter than the gate length. In light of the quantum correction depending on the density change, the modification of the current by the quantum effects is understandable, since the electron density is high and the density change occurs in a short distance. As we expected, similar density distribution across the conduction channel (to what we found in the GaAs/AlGaAs HEMT) is found in this device [7]. The electron densities with and without quantum potential

included are plotted in Fig. 5, which shows the increase of the electron density in the channel when quantum potential is included.

Velocity overshoot, with peak velocity 2.6×10^7 cm/s, was observed in the strained Si channel for both gate lengths used in this simulation, and is very important in achieving the transconductance observed. In Fig. 6, we plot the longitudinal velocity along the conduction channel in the quantum well. The bias condition in this case is $V_g = 0.5$ V and $V_d = 1.5$ V. The velocity overshoot in the gate region results in a peak velocity of 2.6×10^7 cm/s. The overshoot is important in achieving the high transconductance for the device, for it introduces larger current flow along the quantum well. The first velocity peak in the plot is due to the model structure we used for the change of interface discontinuity [7], although it is not practical, it does suggest that the structure can increase the electron velocity between source and gate, which in turn will raise the average velocity through the device and enhance the device performance.

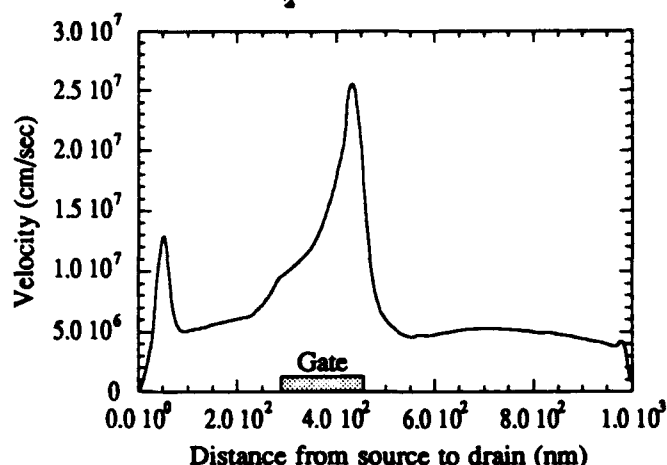


Fig. 6 Longitudinal velocity in the quantum well.

References

- [1] K. Ismail, B. S. Meyerson, S. Rishton, J. Chu, and J. Nocera, "High-Transconductance n-type Si/SiGe modulation doped field effect transistors," *IEEE Electron Dev. Lett.* vol. 13, no. 5, pp. 229-231, May 1992.
- [2] H. Miyata, T. Yamada, and D. K. Ferry, "Electron transport properties of a strained Si layer on a relaxed $\text{Si}_{1-x}\text{Ge}_x$ substrate by Monte Carlo simulation," *Appl. Phys. Lett.* vol. 62, no. 21, pp. 2661-2663, 1993.
- [3] D. Többen, F. Schäffler, A. Zrenner, and G. Abstreiter, "Magnetotransport measurements and low-temperature scattering times of electron gases in high-quality Si/Si_{1-x}Ge_x heterostructures," *Phys. Rev. B* 46, pp. 4344-4347, 1992.
- [4] D. Monroe, Y. H. Xie, E. A. Fitzgerald, and P. J. Silverman, "Quantized hall effects in high-electron-mobility Si/Ge structures," *Phys. Rev. B* 46, pp. 7935-7937, 1992.
- [5] J.-R. Zhou, and D. K. Ferry, "Simulation of ultra-small GaAs MESFET using quantum moment equations," *IEEE Trans. Electron Devices*, vol. 39, No. 3, pp. 473-478, 1992.
- [6] T. J. Bordelon, X.-L. Wang, C. M. Maziar, And A. F. Tasch, "Accounting for bandstructure effects in the hydrodynamic model: a first order approach for silicon device simulation," *Sol.-State Electron.*, vol. 35, no. 2, pp. 131-139, 1992.
- [7] J.-R. Zhou and D. K. Ferry, "Simulation of quantum effects in ultrasmall HEMT devices," *IEEE Trans. Electron Devices*, vol. 40, pp. 421-427, 1993.

Parallel algorithms for the simulation and visualisation of interacting few-electron quantum transport in laterally patterned low-dimensional semiconductors

James Cluckie and John R. Barker

Nanoelectronics Research Centre
Department of Electronics and Electrical Engineering
University of Glasgow, Glasgow G12 8QQ, UK

Abstract

A parallelised Chebyshev accelerated over-relaxation algorithm is developed for simulating the few electron problem in high magnetic fields. A novel form of resonant tunnelling is proposed to occur through a single potential barrier due to the confining potential provided to one electron by a following electron in quasi-one dimensional current flow.

I. Introduction

Arrays of coupled ultra-small capacitor structures exhibit significant charging effects due to the large charging energies deriving from the small values of the components of the capacitance matrix. In particular, the occurrence of single-electronic effects such as the Coulomb blockade, correlated single-electron tunnelling, and the demonstration of switchable single flows, has raised hopes of a future single-electronics technology. The "orthodox model"[1] of correlated single-electron tunnelling has been highly successful for the relatively large metal-insulator systems; it assumes continuous energy distributions, large numbers of states in the 'metallic' electrodes and simplified pictures of tunnelling and electron correlation. However, many tenets of the orthodox theory are lost in the 2DEG semiconductor systems and in the recently proposed coupled Schottky dot structures[2]. Rather than a single electron picture we must deal with a *few-body problem* in which only distant dense 'electrodes' might be treated by self-consistent mean field approaches.

Previously[3,4,5] we have developed a vectorised algorithm for the numerical modelling of 2D quantum transport through quantum point contact structures, quantum waveguides and Aharonov Bohm ring devices using the ADI algorithm on the discretised 2D Schrödinger equation. This method fails for high magnetic fields due to errors arising from non-commutativity of the split kinetic energy operators. It is also unacceptable in accuracy for the few body problem and regions of complex quantum chaos.

II. Algorithms

This work is part of a much more general problem which concerns the modelling of the transport of *n-interacting electrons* within an arbitrary shaped quantum waveguide (hence discrete energy states) including *charging effects*. Our algorithm has been structured such that the interacting 2 electron problem can be studied for a 1D spatial model. The 2 electron problem in configuration space is isomorphic to the problem of 1 electron in 2D space in the presence of an effective 2D potential given by the sum of the confinement potentials for each electron plus the Coulomb potential. Consider first the 1 electron 2D time-dependent Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} \psi(x, y, t) = H\psi(x, y, t) \quad (1)$$

which has solutions of the form

$$\psi(x, y, t) = \psi(x, y, 0) \exp\left\{-\frac{iHt}{\hbar}\right\} \quad (2)$$

or

$$\psi(x, y, t + \Delta t) = \psi(x, y, t) \exp\left\{-\frac{iH\Delta t}{\hbar}\right\} \quad (3)$$

Using the Cayley Expansion for the exponential and defining $\tau = \frac{i\Delta t}{2\hbar}$

$$\psi(x, y)^{t+1} = \psi(x, y)^t \frac{1 - H\tau}{1 + H\tau} \quad (4)$$

Where the t indices refer to discrete time.

Now for a single electron two-dimensional system with time-independent potential profile the time-independent Schrödinger Equation is given by

$$H\psi(x, y) = \left\{-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + V(x, y)\right\}\psi(x, y) \quad (5)$$

Using Taylor Expansions this can be approximated to second order in Δ (the spatial grid spacing) by

$$H\psi'_{x,y} = -\frac{\hbar^2}{2m\Delta^2}(\psi'_{x+1,y} + \psi'_{x-1,y} + \psi'_{x,y+1} + \psi'_{x,y-1} - 4\psi'_{x,y}) + V_{x,y}\psi'_{x,y} \quad (6)$$

Where once again the x and y indices refer to the discrete space. Substituting this in Eqn (4) above, re-ordering and defining various other terms then gives

$$\zeta_{x,y}\psi_{x,y}^{t+1} + \psi_{x+1,y}^{t+1} + \psi_{x-1,y}^{t+1} + \psi_{x,y+1}^{t+1} + \psi_{x,y-1}^{t+1} = \Omega'_{x,y} \quad (7)$$

where

$$\zeta_{x,y} \equiv i\beta - 4 - \gamma V_{x,y} \quad (8a)$$

$$\beta \equiv \frac{4m\Delta^2}{\hbar\Delta t} \quad (8b)$$

$$\gamma \equiv \frac{2m\Delta^2}{\hbar^2} \quad (8c)$$

$$\Omega'_{x,y} \equiv (4 + \gamma V_{x,y} + i\beta)\psi'_{x,y} - \psi'_{x+1,y} - \psi'_{x-1,y} - \psi'_{x,y+1} - \psi'_{x,y-1} \quad (8d)$$

The algorithm we use to solve this set of equations is Chebyshev Accelerated Simultaneous Over Relaxation, a successive approximation method. The residual is defined as

$$\xi_{x,y}^{i+1} \equiv \zeta_{x,y}\psi_{x,y}^i + \psi_{x+1,y}^i + \psi_{x-1,y}^i + \psi_{x,y+1}^i + \psi_{x,y-1}^i - \Omega'_{x,y} \quad (9)$$

where the i index refers to each iteration. This is iterated repeatedly in a chessboard fashion (i.e. odd/even) using

$$\psi_{x,y}^{i+1} = \psi_{x,y}^i - \frac{\omega \xi_{x,y}^{i+1}}{\zeta_{x,y}} \quad (10)$$

until the residual is sufficiently small. The initial wave function for each set of iterations is taken as that for the previous time step. ω is defined by

$$\omega^0 = 1$$

$$\omega^1 = \frac{1}{(1 - \frac{1}{2}\rho_J^2)} \quad (11)$$

$$\omega^{n+1} = \frac{1}{(1 - \frac{1}{2}\rho_J^2\omega^n)}$$

where ρ_J is the spectral radius of the Jacobian method i.e.

$$\rho_J = \frac{1}{2}(\cos(\frac{\pi}{J}) + \cos(\frac{\pi}{L})) \quad (12)$$

and J and L are the number of grid points in each direction.

IV. Inclusion of a magnetic field

The basics of this algorithm remain the same when magnetic field is included and when it is adapted for two 1D interacting electrons. Using the symmetric gauge Eqns (7) & (8) become

$$\zeta_{x,y} \psi_{x,y}^{i+1} + (1 - i\theta y) \psi_{x+1,y}^{i+1} + (1 + i\theta y) \psi_{x-1,y}^{i+1} + (1 - i\theta x) \psi_{x,y+1}^{i+1} + (1 + i\theta x) \psi_{x,y-1}^{i+1} = \Omega_{x,y}^i \quad (7a)$$

where

$$\zeta_{x,y} \equiv i\beta - 4 - \gamma \mathcal{V}_{x,y} - \phi(x^2 + y^2) \quad (8a)$$

$$\beta \equiv \frac{4m\Delta^2}{\hbar\Delta t} \quad (8b)$$

$$\gamma \equiv \frac{2m\Delta^2}{\hbar^2} \quad (8c)$$

$$\Omega_{x,y}^i \equiv (4 + \gamma \mathcal{V}_{x,y} + i\beta + \phi(x^2 + y^2)) \psi_{x,y}^i - (1 - i\theta y) \psi_{x+1,y}^i \quad (8d)$$

$$- (1 + i\theta y) \psi_{x-1,y}^i - (1 - i\theta x) \psi_{x,y+1}^i - (1 + i\theta x) \psi_{x,y-1}^i \quad (8e)$$

$$\theta = meB\Delta^2 \quad (8f)$$

$$\phi = \frac{me^2 B^2 \Delta^4}{2\hbar}$$

The residual is then defined by

$$\xi_{x,y}^{i+1} \equiv \zeta_{x,y} \psi_{x,y}^i + (1 - i\theta y) \psi_{x+1,y}^{i+1} + (1 + i\theta y) \psi_{x-1,y}^{i+1} + (1 - i\theta x) \psi_{x,y+1}^{i+1} + (1 + i\theta x) \psi_{x,y-1}^{i+1} - \Omega_{x,y}^i \quad (9a)$$

and used as before to solve for each successive approximation.

V. Two Electron case

If we consider x and y as the 1D coordinates of two interacting electrons the time independent Schrödinger equation can be written identically to Eqn (5) above with

$$V(x,y) = \frac{e^2}{4\pi\epsilon_0|x-y|} + V_1(x) + V_1(y) \quad (13)$$

where V_1 is the 1D electrostatic confinement potential. Therefore using this as the potential enables us to use identical methods as above.

VI. Parallelisation

The 2 dimensional grid is split evenly over a rectangular array of transputers which iterate their own section of grid and swap boundary conditions at each time step.

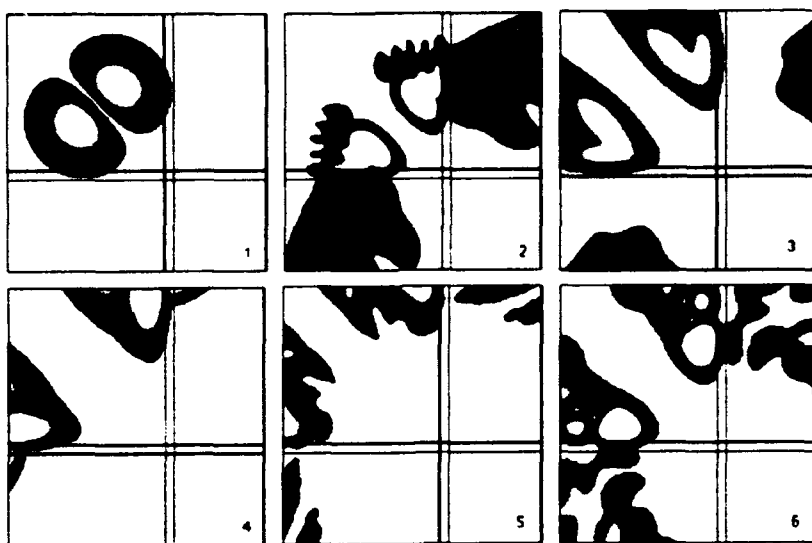


Figure 2(a) Contour plot of modulus of 2-electron wavefunction in configuration space

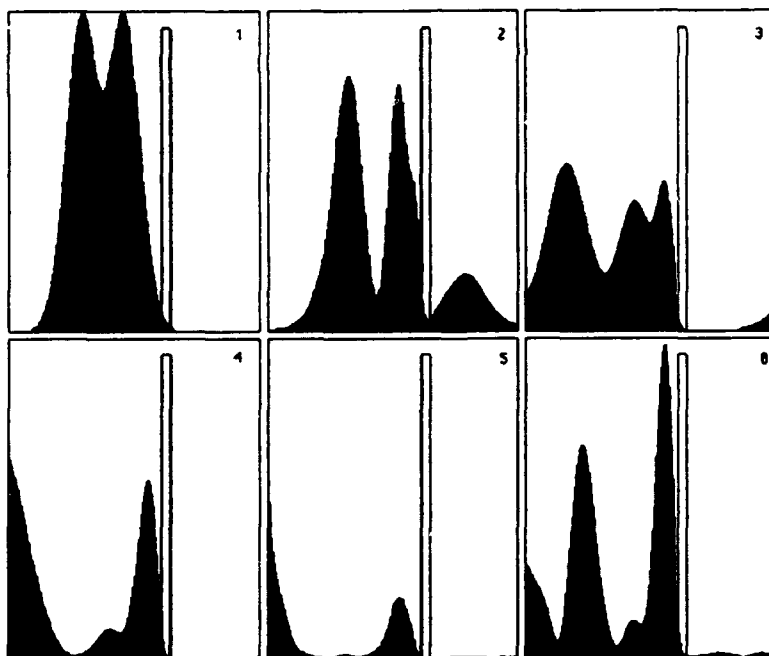


Figure 2(b) Projected total charge distribution in direct space for time sequence of Fig 2a.

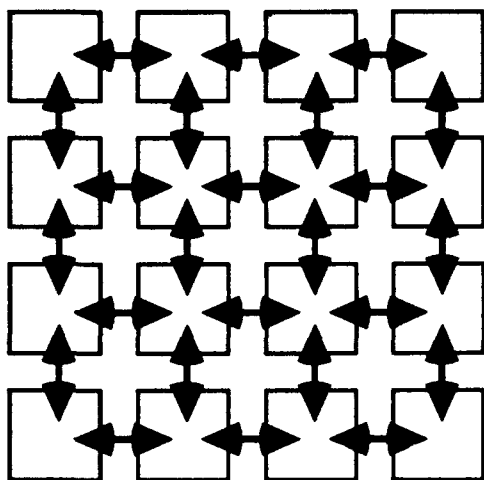


Figure (1) transputer linkage (4 x 4 case)

The method is parallelised on a 64 transputer array and runs at approximately 200 MFLOPS. The results of this study are also being used to validate a novel algorithm for *few electron* transport in 2 dimensions based upon an extension of the *coupled-mode formalism* [8] developed for 1 electron 2D transport. Applications to chaos in magnetic fields are in progress.

VII. Coulomb assisted resonant tunnelling in quasi-1D electron systems.

For the case of 2-interacting electrons transporting/tunneling through bottlenecks in small arrays of quantum point contacts we have identified a new resonant tunnelling process which arises from the tunnelling decay from the quasi-confinement of one electron between a second following electron and a downstream potential barrier. The effect requires mono-mode conduction. Hitherto, only the combination of conventional resonant tunnelling with Coulomb blockade from the charging of the intermediate quantum well has been identified. Figure (2) shows the case of two equal spin electrons initially 30 nm apart (Coulomb energy 4.5 meV) encountering a barrier of height 50 meV, width 5 nm for incident kinetic energy of 10 meV/electron for electrons in GaAs. The effect is strongly damped unless the electron stream is strongly correlated in space and time. Coupled quantum dots, long chain polymers, enzymes, redox chains provide possible test cases.

References

- [1] K K Likhaev, IBM J R & D, 32 144 (1988)
- [2] J R Barker, J Weaver, S Babiker, S Roy, Symposium on New Phenomena in Mesoscopic Structures, Hawaii (1992).
- [3] J R Barker, Chapter 13 of *Semiconductor Device Modelling*, ed. C. M. Snowden, Springer-Verlag: London, 207-226 (1989).
- [4] J R Barker, in *Physics and fabrication of nanostructures*, ed M Read and W. P. Kirk, Academic Press 253 -262 (1989)
- [5] J. R. Barker, M. Finch, J. Pepin and M. Laughton
Solid State Electronics 32 1155-1159 (1989)
- [6] M Laughton, J R Barker, J H Davies, J Nixon, *Phys.Rev B* 44 1150 (1991)
- [7] J R Barker, in *Handbook on Semiconductors* edited by T Moss volume 1 edited by P T Landsberg, Elsevier Science Publishers Chapter 19 1079-1127 (1992).
- [8] J Cluckie and J R Barker, Two-body quantum transport theory of interacting electrons in semiconductor 2DEG structures, 8th International Conference on Hot carriers in semiconductors, Oxford (1993).

System simulation tools for single-electronic devices

Scott Roy, John R Barker and Asen Asenov

Nanoelectronics Research Centre
Department of Electronics and Electrical Engineering
University of Glasgow
Glasgow G12 8QQ, UK

Abstract

Single-electronic systems are described by coupled sets of circuit equations which link the charge, voltage and current distributions of tunnelling and non-tunnelling capacitor arrays. The concept of critical charge is used to implement an efficient network solver. System simulation may then be achieved by Monte Carlo methods although this is often prohibitive computationally. Alternatively, linear programme techniques can establish the boundaries for stable operation. The full 3D modelling of the capacitance matrix is required for recently developed Schottky dot structures.

I. Introduction

The recent development of single-electronic devices has exploited the phenomenon of correlated single electron tunnelling[1] in coupled tunnel junctions (ultra-small capacitor arrays) under conditions set by the Coulomb blockade threshold $e^2/2C \gg kT$. At least three different classes of structure are currently under study experimentally: vertical metal-insulator-metal[2-4], lateral metal semiconductor metal[5]; and laterally patterned two-dimensional electron gases in semiconductor heterostructures[6]. These new devices pose new challenges for computational electronics[7]: they are strongly capacitively coupled, the major transport mechanism is correlated tunnelling, the presence of thermal fluctuations, cross-talk, charge trapping de-trapping and macroscopic quantum tunnelling are all potentially killer effects. Most significantly existing single-electronic systems may be quite large (up to 100 devices) requiring new systems tools for design and analysis.

II. Network solver

We have developed a set of simulation tools based on a similar formalism to Bakhvalov et al[10] for a linear array of tunnel junctions but extended to arbitrary single electron tunnel junction circuit configurations using a matrix representation of the various voltages, currents and circuit elements. The general theory derives from an analysis of a basic tunnelling event and a single tunnel junction (capacitance C) in series with a non-tunnelling capacitance C_e and a voltage source V_0 . The tunnelling rate Γ is determined by the temperature T and the free energy change following the tunnel event[1]: $\Gamma(\Delta E, T) = (\Delta E/e^2 R_t) [\exp(\Delta E/k_B T) - 1]^{-1}$ where R_t is the junction tunnel resistance. At low temperatures

$$\Gamma(\Delta E) = (-\Delta E/e^2 R_t) (\Delta E < 0); \Gamma(\Delta E) = 0 (\Delta E > 0). \quad (1)$$

For the simple circuit we have: $E = Q_e^2/(2C_e) + Q_c^2/(2C) + Q_v V_0$; where Q_v is the charge through the voltage source and the critical charge $Q_c = e/(2[1+C_e/C])$. For a single electron tunnel event we find

$$\Delta E = -eQ/C + e^2/(2(C+C_e)) \quad (2)$$

This "instantaneous" model for tunnelling provides a first level for our system simulation tools. A second level utilises the quantum Langevin equation[] approach handle fluctuations (due to lack of space we do not describe that here). The critical charge is the charge at which tunnelling becomes advantageous; the critical charges for a circuit depend solely on the junction

capacitance and the lumped capacitance C_{eff} of the remaining circuit. We have developed a network solver for C_{eff} by deploying an impedance matrix Z , positive definite with rank equal to the number of loops in the circuit. The diagonal elements are the total impedances around each mesh; the off-diagonal terms are the total impedances shared by two loops. By partitioning this matrix we can separate out loops of no interest to obtain a lower rank effective impedance matrix Z_{eff} . Hence we may determine the critical charges.

III. Monte Carlo and Linear Programming simulators

The Monte Carlo simulator uses the general network solver to find the critical charges for each device for which tunnelling becomes feasible in the tunnel junction network. The simulator then iterates "events" by repeatedly characterising the circuit for a given set of input voltages and clocked charge positions, discovering which tunnelling event will occur next and updating dependent and independent circuit parameters using charge conservation and the circuit matrix equations. Although this approach is important it is computationally expensive especially for the larger extended systems of coupled devices.

To obtain a more rapid assessment of the possible stable operating regimes of single electronic devices and systems we have developed a *linear programming* technique which allows us to determine the allowed regime of stable operation in the control parameter space. The method is essentially an inverse of the Monte Carlo approach: the allowed or disallowed tunnel events are defined first followed by a determination of the circuit voltages and charge values. The approach is illustrated in figures 2-4 for a turnstile device[4] shown schematically in figure 1. The Monte Carlo results for the operational area of V_G - V_A space (fig 4) are given by accumulating legal points. The area so-defined is found to be well-modelled by boundary lines determined from linear programming using the tunnel event schematics of figure 3.

These new simulation tools have been used to study effects of cross-talk and inter array coupling in single-electronic systems. Figure 6 shows typical results of the effects of stray capacitance coupling C_{stray} on the current through two parallel tunnel junction arrays shown in figure 5.

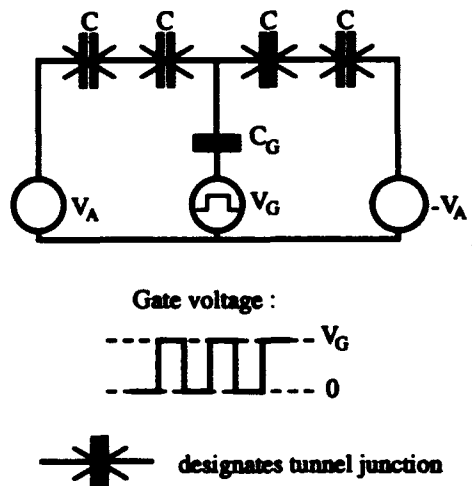


Figure1. 2-phase turnstile device

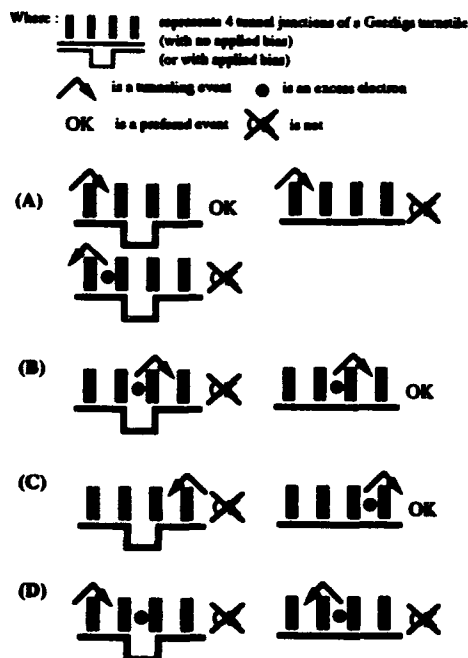


Figure2. Linear programming schematics

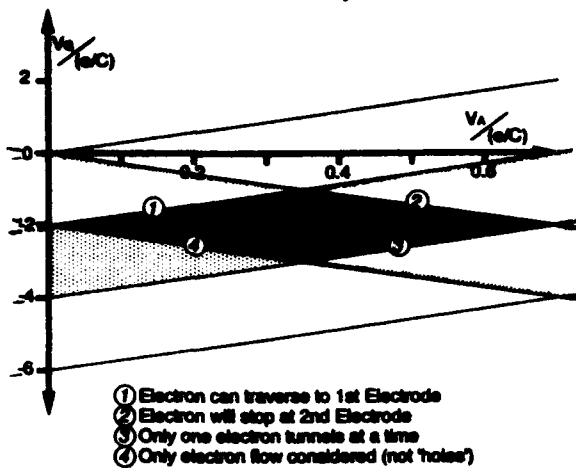


Figure 3. Legal area of turnstiling operation in VG, VA space. Dark shading = safe operation

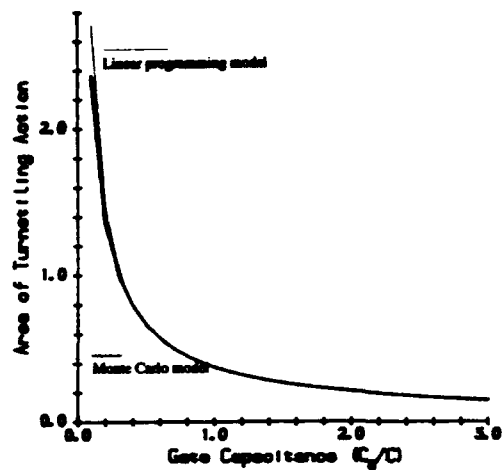


Figure 4. Legal area of turnstiling operation in VG, VA space versus gate capacitance

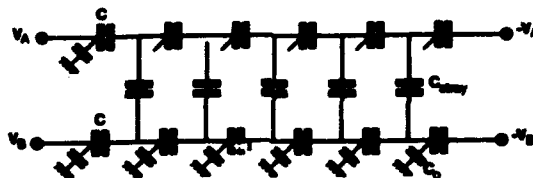


Figure 5. Linked tunnel junction arrays

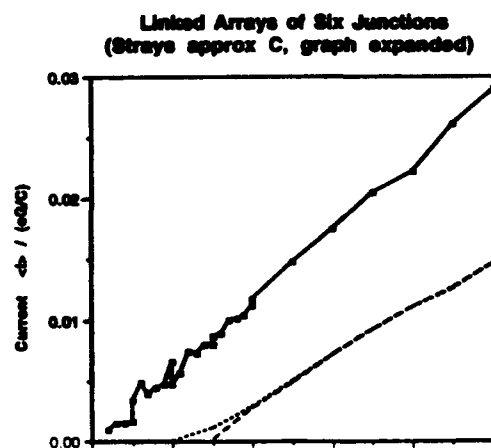
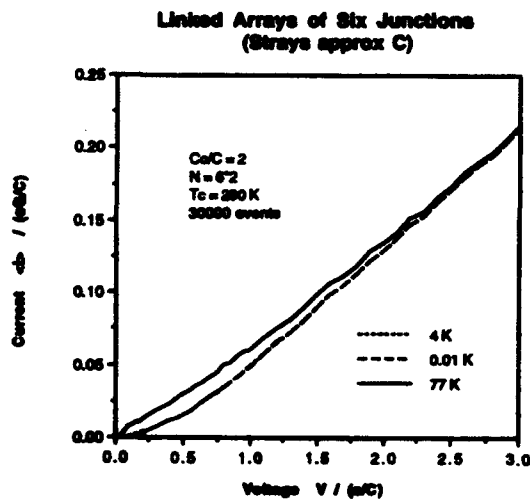


Figure 6. Current-voltage characteristics for Linked tunnel junction arrays

IV. Schottky dot systems

The traditional approach to single-electronic devices has utilised the hanging-resist technology developed for metal-insulator-metal tunnel junctions[2]. Although this method has allowed construction of large numbers of interconnected junctions in complex circuits it is limited by the lithography to relatively large capacitances and consequently very low operating temperatures. Very recently a much finer resolution lithography which has fewer processing stages and involves forming ultra-fine (< 5 nm radii and spacings) metal on semiconductor electrodes, Schottky islands and dot arrays has been developed at Glasgow[5](Fig 7). These new structures involve ultra-small capacitances with equivalent Coulomb blockade temperatures in excess of 60K scaleable to much greater than room temperature. The dot structures may be arranged laterally to form pass transistors (fig8), RAM cells (fig9) or more complex circuits. An essential feature of these systems is the requirement for the "tunnelling tails" of the Schottky islands to overlap thus permitting correlated electron tunnelling from dot to dot via the semiconductor. Modelling of these systems is crucial in order to determine the effective capacitance matrix and for understanding how to control effects of unwanted traps (charge-trapping de-trapping effects) which can destroy single electronic stability by structuring the substrate, ground planes, doping levels and island geometry.

The development of stable and reliable single electronic systems requires the precise design of both the junction capacitances (inter-capacitances) and the capacitances to ground. Since the measurement of such ultra-small capacitances is very difficult experimentally the design becomes reliant on the numerical simulation of the capacitance matrices. The problem is essentially a 3D problem involving complicated geometries, several dielectric regimes, and device physics which involves surface conditions, random distributions of traps and their dynamics.

To illustrate part of the design problem and the importance of an adequate numerical solution we present here some results of 2D simulation of Aluminium wires on the surface of p-silicon as an approximation to the Schottky dot devices shown in Figures 7-9. The simulations were performed by the simulator H2F [8]. Figure 10 shows the potential distribution around two 40 nm width wires with 12 nm spacing. The fringing effects and presence of the silicon substrate increased the junction capacitance 4 times in comparison with the simple parallel plate formula. The ground capacitance is more than twice as high as the junction capacitance. The presence of surface pinning states modifies the picture. The presence of donor type states near the middle of the bandgap increases the ground capacitance slightly, but a more profound effect is produced by modelling acceptor states which significantly increase the capacitance(25%). The junction and ground capacitances may be tuned by anisotropic etching leaving metal islands on the top of silicon pedestals (figure 11). It is found that 200nm etching reduces the junction capacitance by a factor of 3x and the capacitance to ground is diminished by a factor of 2x. A parallelised 3D simulator is under development for more realistic prediction and design of such devices[9].

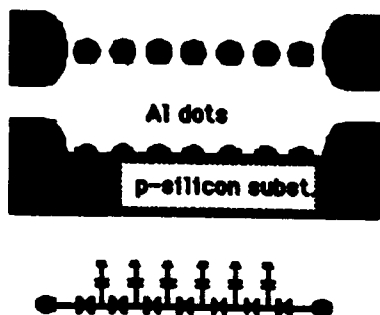


Figure 7. Schematic and equivalent circuit of Schottky dot arrays

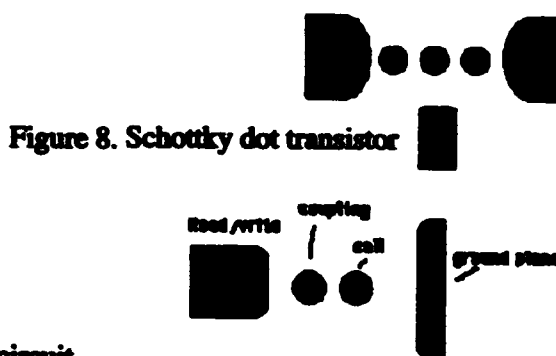


Figure 8. Schottky dot transistor

Figure 9 Schottky dot RAM cell

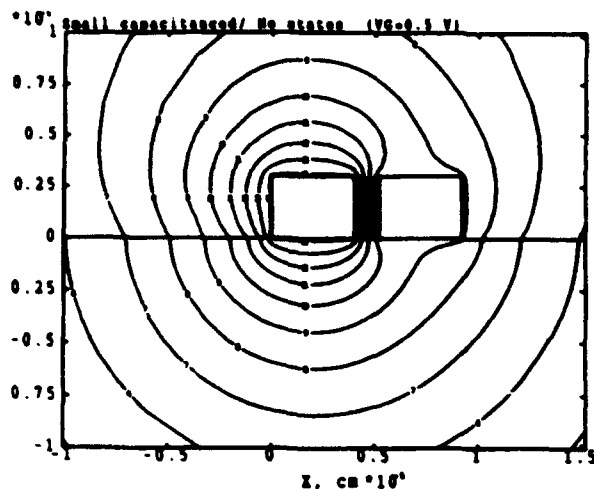


Figure 10. Potential distribution around two Schottky dots on Si.

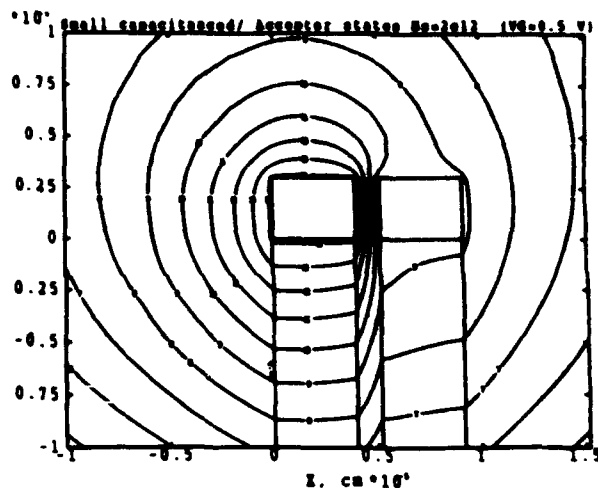


Figure 11. Potential distribution around two Schottky pedestals on Si. $V_G = 0.5V$

References

- [1] D V Averin and K K Likhaev, J Low Temp Phys 62 345 (1986).
- [2] T A Fulton and G J Dolan, Phys Rev Letters, 89 109 (1987).
- [3] L S Kuzmin, P Delsing, T Claesson, K K Likhaev, Phys Rev Letters 60 309 (1989), 62 2539 (1989)
- [4] L J Geerligs, V F Anderagg, P A M Holweg, J E Mooij, H Pothier, D Esteves, C Urbina, M H Devoret, Phys. Rev. Letters 64 2691 (1990).
- [5] J R Barker, J M R Weaver, S Babiker and S Roy, *Theory, modelling and construction of single-electronic systems*, Proceedings Second International Symposium on New Phenomena in Mesoscopic Structures, Hawaii, (1992).
- [6] U Meir, M A Kastner and S J Wind Phys RevB40 5871 (1989).
- [7] R Barker, S Roy and S Babiker, *Trajectory representations, fluctuations and stability of granular electronic devices*, Chapter 22, pp213-232, Science and Technology of Mesoscopic Structures, edited by S. Namba, C. Hamaguchi, T. Ando. Springer-Verlag: London, Tokyo, New York. (1992)
- [8] A. Asenov, D. Reid, J.R. Barker, N. Cameron and S. P. Beaumont, *Finite element simulation of recess gate MESFETs and HEMTs. The simulator H2F*, Accepted for SISDEP, Wien, Austria (1993).
- [9] A. Brown, D. Reid, A. Asenov and J. R. Barker, *The implementation and speed-up of coloured SOR methods solving 3D Poisson's equation on arrays of transputers*. Submitted for this Workshop (1993).
- [10] N S Bakhvalov, G S Kazach, K K Likhaev, S I Serdyukova, Sov. Phys. JETP 68 581 (1989)
- [11] K K Likhaev, IBM J Res Dev 32 144 (1988)
- [12] D Esteve in lectures by L J Geerligs and J. E Mooij, *Granular Nanoelectronics*, ed D K Ferry, J R Barker and C Jacoboni, Plenum Press, (1990).

Adaptive potential in alternating-direction implicit relaxation

Nam Young Kim and A. M. Kriman

Dept. of Elec. and Comp. Engineering,
State University of New York at Buffalo,
Buffalo, New York 14260 USA

Abstract

We demonstrate a useful decomposition of the potential energy, which accelerates the relaxation method for finding the ground state of a Schrödinger operator in multiple space dimensions, and improves the ultimate accuracy achievable. The potential decomposition makes the Schrödinger operator approximately separable for the evolving approximation to the ground state; this decreases the error associated with the long time-steps in alternating-difference implicit schemes.

I. Review and Motivation

A central step in many quantum modeling problems is to find the lowest-energy eigenstates of a Hamiltonian operator. A standard approach to find them is the relaxation technique [1].

The relaxation technique, applied to find the ground state of Hamiltonian H , amounts to evolution of the Schrödinger equation in imaginary time:

$$H\varphi(t) = -\hbar \partial_t \varphi(t), \quad (1)$$

Starting from an initial condition that has nonzero overlap with the ground state, the ground state is asymptotically dominant after imaginary times long compared to $\tau_{10} \equiv \hbar / (E_1 - E_0)$,

$$\varphi(t) \sim a_0(t) \varphi_0. \quad (2)$$

(E_0 and E_1 are the ground and first excited energies of the Hamiltonian, respectively; a_0 is a space-independent coefficient.) The primary goals in such an evolution are stability and accuracy. Implicit techniques exist which are well-known to yield stable evolution both in real and imaginary time, in which accuracy can always be assured by using a sufficiently small time step Δt . The simplest of these is

$$(1 + \Delta t H / \hbar) \varphi^j = \varphi^{j-1}, \quad (3)$$

with $\varphi^j \equiv \varphi(t_j)$; $t_j = j\Delta t$.

However, just as not all eigenstates are required, similarly not every kind of accuracy is required either. In particular, since the object of the relaxation is simply to remove the high-energy components as rapidly as possible, in order to examine what is left, an accurately exponential decay of those components is less important than their mere rapid disappearance.

For this reason, it is common to accelerate the relaxation procedure by using long time steps. When this acceleration is used with one-dimensional Hamiltonians, one converges to the ground state.

The evolution described in (3), like other implicit evolution methods, requires inversion of a large matrix (in this case, the matrix representing $1 + \Delta t H / \hbar$). The matrix to be inverted has a dimension equal to the number mesh points. This dimension grows as the inverse of the mesh

spacing raised to the power of the mesh dimension. For multidimensional problems, this inversion is impractical or numerically intractable. Instead, standard time-evolution techniques for multi-dimensional Hamiltonians typically use an operator-separation or alternating-direction implicit (ADI) scheme. For example, decomposing a two-dimensional Hamiltonian H^{2D} as

$$H^{2D} = H^x + H^y, \quad (4)$$

one may write

$$\phi^j = \frac{1}{1 + \Delta t H^x/\hbar} \frac{1}{1 + \Delta t H^y/\hbar} \phi^{j-1}, \quad (5)$$

or equivalently:

$$(1 + \Delta t H^y/\hbar) \phi^{j-0.5} = \phi^{j-1}, \quad (6a)$$

$$(1 + \Delta t H^x/\hbar) \phi^j = \phi^{j-0.5}. \quad (6b)$$

If H^x and H^y involve only the kinetic energy of motion along x and y directions respectively, then each step (6a, 6b) requires the solution only of a tridiagonal matrix.

In ADI schemes, short time steps are necessary not only for the accurate exponential decay of high-energy states, *but also for the accurate estimation of the ground state*. For example, in (5), ϕ approaches the ground state of

$$H' \equiv H^{2D} + H^x H^y \frac{\Delta t}{\hbar}, \quad (7)$$

which by perturbation theory implies a ground state energy that also is accurate only to first order in Δt . As a result, while short time steps may be used initially to accelerate the dissipation of high-energy components, eventually short time steps must be used to achieve accurate ground states. In this kind of variable time-step relaxation, it can be difficult to determine unambiguously how well one is converging. (See however, work by Doss and Miller on ADI solution of Laplace and related equations [2], in which a way is found to optimize the time step.) In general, efficient use of this approach can require a certain amount of user interaction, and is correspondingly difficult to program for systematic application.

Separable Hamiltonians constitute an important exception to the above distinction between one-dimensional and higher-dimensional. That is, if an ADI scheme consists of the alternating application of the separated components of a separable Hamiltonian, then relaxation converges to the appropriate ground state even for large time steps. [This can be seen immediately in (7), from the fact that the ground state is simultaneously an eigenstate of H^x and H^y as well as H^{2D} .] Most Hamiltonians (almost all, in the appropriate probability-measure sense) cannot be put into separable form. However, as we describe below, it is possible to find a decomposition that is separable in a restricted sense: the Hamiltonian is partitioned in such a way that at least the ground state is simultaneously an eigenstate of each partitioned component. This permits an acceleration of relaxation without the usual penalty in accuracy.

II. Adaptive Potential

Most Hamiltonians (almost all, in the appropriate probability-measure sense) cannot be put into separable form. However, as we describe below, it is possible to find a decomposition that is separable in a restricted sense, thus permitting an acceleration of relaxation without the usual penalty in accuracy.

In order to simplify the presentation, we here consider only two-dimensional Hamiltonians (4). We seek new potentials $V_0^x(x,y)$ and $V_0^y(x,y)$ so that, with

$$H_0^x = K^x + V_0^x, \quad H_0^y = K^y + V_0^y, \quad (8)$$

the ground state ϕ_0 of H^{2D} also satisfies the auxiliary equations

$$H_0^x \phi_0 = H_0^y \phi_0 = \frac{1}{2} E_0 \phi_0. \quad (9)$$

If ϕ_0 is known, (9) can be solved for the potentials using (8). Because the total potential must always be V , it suffices to determine only the single function

$$\Delta V \equiv \frac{1}{2} (V^x - V^y). \quad (10)$$

During the time evolution, one knows only ϕ , the evolving best approximation to ϕ_0 . To indicate this we omit the zero subscripts on the separated potentials. Assuming that $\phi - \phi_0$ small, we "solve" (9) to find

$$\Delta V[\phi] = - \frac{(K^x - K^y)\phi}{2\phi} \quad (11)$$

In the adaptive-potential ADI (APADI) approach, we recompute ΔV after each time step. Using this ΔV directly to define the current potential is unstable, so instead we relax exponentially toward the instantaneous potential, with ΔV for the i th time step defined by:

$$\Delta V^i \equiv (1-\alpha)\Delta V^{i-1} + \alpha \Delta V[\phi^{i-1}], \quad (12)$$

and depending on a relaxation parameter α .

III. Numerical Results

We have applied the potential (12) in both first-order and second-order operator separation schemes. We first treat a separable example (a sum of finite square-well potentials, in the x and y directions) in order to compare the adaptive-potential ADI (APADI) scheme not only with naïve-potential ADI (NPADI: $V^x = V^y = V/2$) but also with the a partition using the exact separated potentials ADI (SPADI). Figure 1 compares these in a first-order scheme. We plot the fractional deviation of the computed ground state energy from the exact ground-state energy, using a constant time step ($\Delta t = 2$, in units chosen so that the electron mass, ångström, and electron-volt have unit magnitude), with a one-dimensional well depth of 0.2 eV, a well-width of 40 Å, and electron mass of $0.1 \times$ free-electron mass. The relaxation parameter α is 0.05. APADI clearly represents a significant improvement over the standard approach, represented by NPADI.

After $t \approx 200$, SPADI fluctuates due to round-off error. Using second-order operator separation ("true" ADI), the APADI scheme is significantly better, to the extent that APADI and SPADI differ only in this round-off error regime.

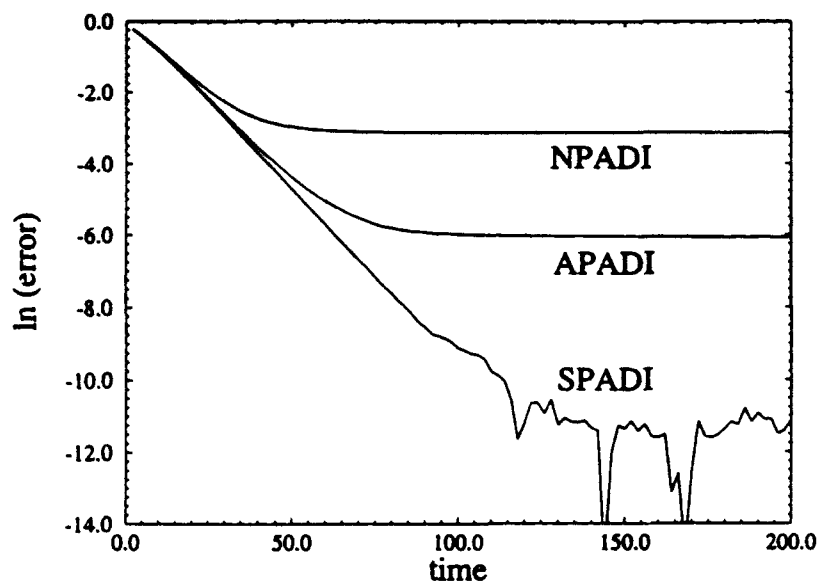


Figure 1. Fractional error in the ground state energy, found using first-order operator separation for an exactly separable problem.

A nonseparable example, illustrated below, is based on the total potential illustrated in figure 2 (piecewise constant: 0.0 eV in first and third quadrants, 0.2 eV in the remaining two, with hard walls around a square of side 400Å).

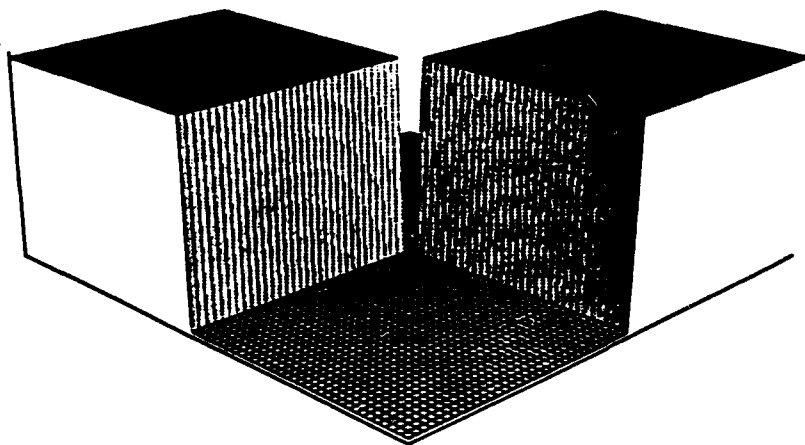


Figure 2. A nonseparable potential, with the x axis along the right-hand edge, and the y axis along the left-hand edge.

After relaxing toward a solution, V^x had the form shown in Fig. 3. Note that V^x is different for each energy level. The special case in which it is the same for every level is the usual case of a separable potential.

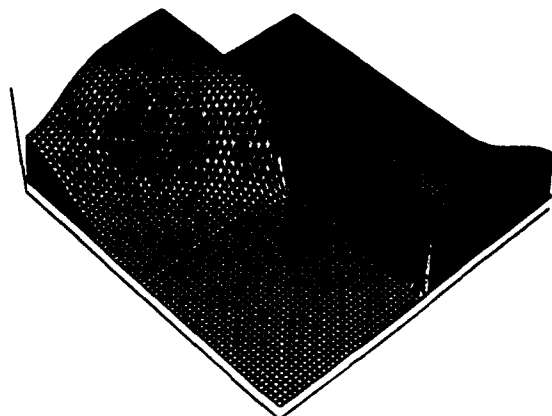


Figure 3. V^x for ground state of V in figure 2 (different scale, perspective; x-axis still to right)

The graph corresponding to Fig. 1, for the nonseparable potential, is Fig. 4. This calculation was done using a second-order ADI.

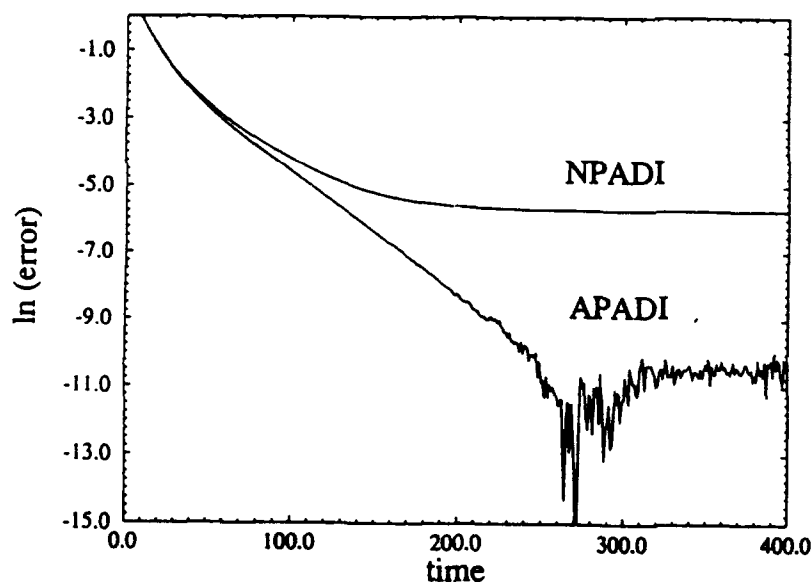


Figure 4. Fractional error for nonseparable potential.

References

- [1] See, for example, W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in FORTRAN*, Cambridge Univ. Pr., 1986, 1992.
- [2] S. Doss and K. Miller, "Dynamic ADI Methods for Elliptic Equations," *SIAM J. Numer. Anal.*, Vol. 16, No. 5, pp. 837-856, October 1979.

Numerical study of 2-D quantum dots

M. Macucci

Dipartimento di Ingegneria dell'Informazione,
Università di Pisa,
Via Diotisalvi, 2 I-56126 Pisa, Italy

Karl Hess

Beckman Institute,
University of Illinois at Urbana-Champaign,
405 N. Mathews, Urbana, Illinois 61801

Abstract

We present a numerical study of the chemical potential and of the capacitance in a model quantum dot. Our model includes the electron-electron interaction and exchange and correlation effects within the framework of density functional theory. Our results exhibit the typical features observed in experiments, such as the increase in the capacitance for increasing number of electrons and the presence of irregularities in the succession of the chemical potential values vs. the electron number.

I. Introduction

Recent experimental results have shown extremely interesting single-electron phenomena in semiconductor quantum dots. The conductance across a quantum dot loosely coupled to the external electrodes via low-transparency tunnel barriers has been shown to be periodic with respect to the voltage applied to suitably positioned gates [1-3]. This effect has been explained as the consequence of the chemical potential in the dot lining up with the one in the external electrodes [2]. If the electrostatic energy prevails over the quantum confinement energy, the behavior of the dot is substantially capacitive and we expect evenly spaced values of the chemical potential versus the number of electrons. Several theoretical studies of quantum dots exist in the literature [4-5], dealing with both very idealized models and with realistic, three-dimensional models. Our aim has been to solve for the chemical potential in a structure which, even though simplified, retains all the characteristic features observed in the experiments. We study a 2-D model quantum dot defined by a given confining potential. Except for this simplification, i.e. the confining potential not determined from the actual characteristics of the semiconductor layers and the geometries and voltages of the metal gates, we try to take into account all relevant contributions, including many-body effects, within the framework of density functional theory.

II. Quantum dot model

Our model quantum dot is two-dimensional and defined by the shape of the confining potential, represented by hard walls along the perimeter of a square region to which we shall refer in the following as "quantum box". The potential within the box is not constant, but quasi-parabolic, as the one produced by a positive background charge of $100q$ (q being the electron charge), uniformly distributed over the surface of the box.

The potential effectively seen by the electrons in the calculations presented in this paper is substantially the quasi-parabolic part, since this already provides enough confinement as to make the electron density vanish before reaching the hard walls. This is a realistic approximation of the actual potential in the experimental situations.

Our study is performed at the temperature of 0 K, thus for a system in the ground state. We solve for the eigenfunctions and the eigenenergies of the quantum dot by means of an iterative, self-consistent procedure [8] based on the Kohn-Sham density functional approach [9].

The Schrödinger equation is discretized with a standard five-point formula and the eigenvalues and eigenvectors of the matrix thus obtained are computed by means of a Ritz iterative procedure.

III. Numerical results

The chemical potential is the main quantity we want to evaluate. Once the wave functions and the energy eigenvalues have been determined, there are several ways to obtain the chemical potential $\mu(N)$ (N being the number of electrons). We can compute the total energy $E(N)$ of the system from [9]

$$E(N) = \sum_{i=1}^N \varepsilon_i - \frac{1}{2} \int \int \frac{n(\vec{r})n(\vec{\rho})}{|\vec{r} - \vec{\rho}|} d\vec{r}d\vec{\rho} + \int n(\vec{r}) [E_{ex}(n(\vec{r})) + E_{corr}(n(\vec{r})) - V_{ex}(n(\vec{r})) - V_{corr}(n(\vec{r}))] d\vec{r}, \quad (1)$$

where ε_i are the energy eigenvalues for each electron, $n(r)$ is the total electron density, E_{ex} , E_{corr} , V_{ex} , V_{corr} are the exchange and correlation energies and potentials, respectively.

From the definition of chemical potential we have that $\mu(N) = E(N) - E(N - 1)$. The result of this differentiation, however, may be adversely affected by numerical errors in the values of $E(N)$ and $E(N - 1)$. A plot of μ versus N obtained with this procedure for a 60×40 nm box is shown in Fig. 1 with solid dots. In the same figure we report, with empty squares, the chemical potential computed for the same structure with a different method: Slater's approximation. The "removal energy", i.e. the energy needed to remove one electron from a system of interacting electrons can be approximated [10] by $\varepsilon(N + 0.5)$, the energy eigenvalue for a fictitious additional particle with charge $q/2$. It is apparent that for a small number of electrons both methods yield the same result while, for more than 10 electrons, the technique based on the difference between the total energies for N and $N - 1$ electrons starts being severely affected by numerical noise. We have therefore used Slater's approximation in most of our calculations.

In Fig. 2 we report an idealized representation of the conductance peaks which would be measured between two leads very loosely coupled through the a square quantum box of various sizes. The height of the peaks is purely conventional and there is no broadening because of the assumption of 0 K temperature and of vanishingly small transparency for the tunnel barriers. The three plots correspond to three different box sizes: 40×40 nm, 80×80 nm and 200×200 nm.

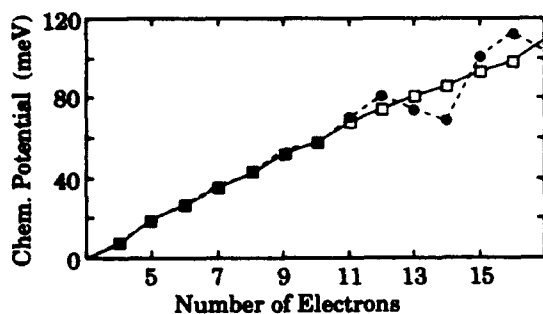


Fig. 1 Chemical potential vs. number of electrons

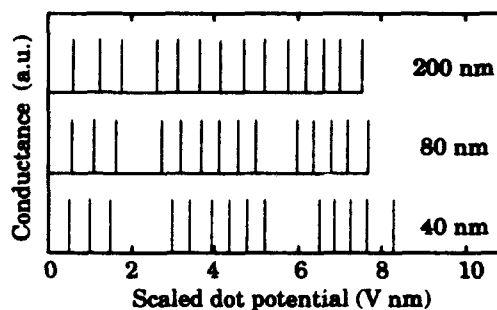


Fig. 2 Conductance peaks for a square quantum box

The quantity on the abscissae axis is the dot potential scaled multiplying it by the length of the box side in nanometers. This scaling normalizes the three plots with respect to the energy associated with the Coulomb interaction [1]. The first peak, coincident with the zero reference for the dot potential, is for three electrons.

The spacing between conductance peaks is very uneven for the smallest box and a more careful exam of the plot shows that there are groups of evenly spaced peaks corresponding to the various degenerate energy levels in the square box. For example, the peaks for the 3rd, 4th, 5th and 6th electrons correspond to the 2nd and 3rd single-electron orbitals, which are degenerate in a square box, in the absence of electron-electron interaction.

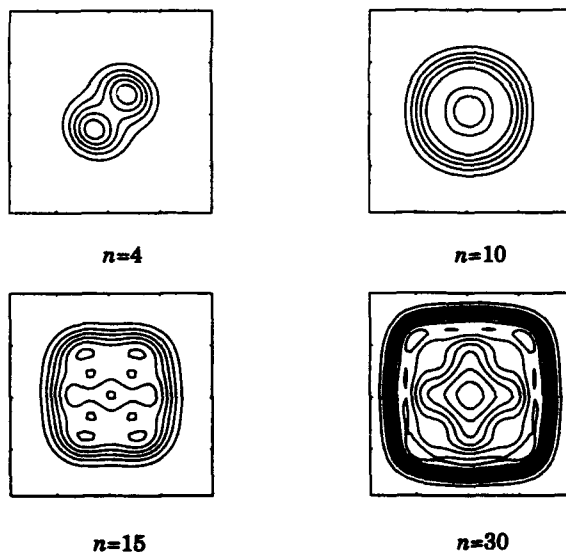


Fig. 3 Electron density in a square quantum box.

As the size of the dot increases, we observe a reduction of the relative importance of the quantization energy and the clear prevalence of the Coulomb energy, leading to an almost even spacing for the 200×200 nm box. The plot for this largest box also shows an effect which has been experimentally observed: the reduction of the average spacing between peaks with increasing number of electrons. This effect is mainly due to the increase of the area of significant electron density when more electrons are added to the

dot and see a less attractive potential, because of the screening performed by the ones in the lower orbitals.

In Fig. 3 contour plots of the electron density are shown for a number of 4, 10, 15, 30 electrons confined in a 120×120 nm quantum box. There is a very significant change in the extension of the non-zero electron density and in its shape when the number of electrons is increased. This accounts for the observed crowding of conductance peaks. A differential capacitance C_d can be associated with a quantum dot, according to the following definition:

$$C_d(N) = \frac{q^2}{\mu(N+1) - \mu(N)}. \quad (2)$$

This quantity is more readily understandable if we consider a dot in the proximity of some conducting surface such as, for example, a metal gate. This is also the most common configuration in the devices which have been experimentally investigated. The differential capacitance represents the ratio of the electron charge to the variation of the voltage between the gate and the dot when an electron is added to the system. If the gate is much larger than the dot, it can be approximated with an infinite conducting surface and modelled with properly placed image charges.

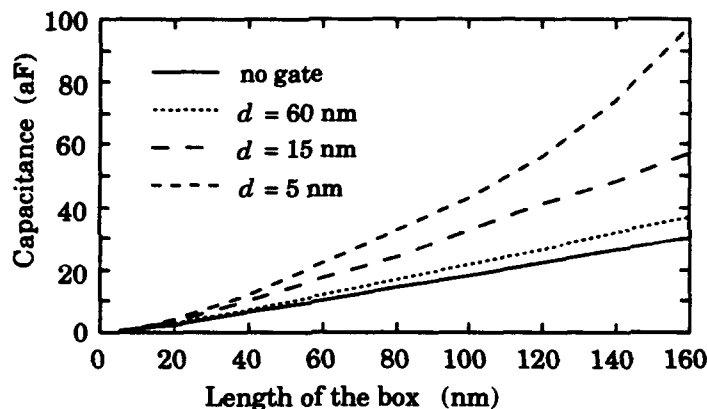


Fig. 4 Capacitance vs. length for 4 electrons.

The results of our calculations of the capacitance [8] in the presence of metal gates are shown in Fig. 4. The four curves are for 4 electrons in a rectangular quantum box with a length/width ratio of 4/3 and a metal gate at a distance d of 5, 15, 60 nm or no gate at all. We see that, while for no gate and for a distant gate the capacitance grows in a substantially linear fashion as we would expect for a 2-D geometry isolated in 3-D space (see e.g. [2]), for a very close gate the behavior becomes almost quadratic, reaching the well-known classical limit of the parallel plate capacitor. For a small gate-dot separation we perform simply a Hartree calculation, because the expressions used for the exchange and correlation potentials [8] do not hold in the presence of strong interaction with the image charges.

IV. Conclusion

We have studied a model two-dimensional quantum dot with a quasi-parabolic confining potential, including many-body effects within a density functional approach. Results for

the chemical potential in a square box have been presented, showing the transition from a behavior dominated by quantum effects to one in which the Coulomb charging energy is predominant. The differential capacitance which can be thus defined approximates the one of a conducting 2-D surface with a shape like the one of the area of significant electron density in the quantum dot.

Acknowledgments

We wish to thank G. J. Iafrate for valuable discussions.

This work was supported by the U.S. Army Research Office, the Office of Naval Research and the National Center for Computational Electronics.

One of us (M.M.) acknowledges also support from the National Research Council of Italy (CNR), and from the Italian Ministry of University and of Scientific and Technological Research.

References

- [1] U. Meirav, M. A. Kastner, and S. J. Wind, "Single-Electron Charging and Periodic Conductance Resonances in GaAs Nanostructures," *Phys. Rev. Lett.*, vol. 65, pp. 771-774, Aug. 1990.
- [2] L. P. Kouwenhoven, N. C. van der Vaart, A. T. Johnson, W. Kool, C. J. P. M. Harmans, J. G. Williamson, A. A. M. Staring, and C. T. Foxon, "Single electron charging effects in semiconductor quantum dots," *Z. Phys. B*, vol. 85, pp. 367-373, 1991.
- [3] H. Pothier, J. Weis, R. J. Haug, K. von Klitzing, K. Ploog, "Realization of an in-plane-gate single-electron transistor," *Appl. Phys. Lett.*, vol. 62, pp. 3174-3176, June 1993.
- [4] G. W. Bryant, "Electronic Structure of Ultrasmall Quantum-Well Boxes," *Phys. Rev. Lett.*, vol. 59, pp. 1140-1143, Sept. 1987.
- [5] A. Kumar, S. E. Laux, and F. Stern, "Electron states in a GaAs quantum dot in a magnetic field," *Phys. Rev. B*, vol. 42, pp. 5166-5175, Sept. 1990.
- [6] W. Häusler, B. Kramer, J. Masek, "The influence of Coulomb interaction on transport through mesoscopic two-barrier structures," *Z. Phys. B*, vol. 85, p. 435, 1991.
- [7] M. Stopa, "Quantum dot tunneling coefficients and capacitances in a magnetic field, Coulomb blockade regime: self-consistent computation," in *Proc. 2nd Intern. Symposium on "New Phenomena in Mesoscopic Structures"*, Maui, Hawaii, USA, pp. 174-177, Dec. 1992.
- [8] M. Macucci, Karl Hess, and G. J. Iafrate, "Electronic Energy Spectrum and the Concept of Capacitance in Quantum Dots," to be published.
- [9] W. Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," *Phys. Rev.*, vol. 140, pp. A1133-A1138, Nov. 1965.
- [10] J. P. Perdew and Alex Zunger, "Self-interaction correction to density-functional approximations for many-electron systems," *Phys. Rev. B*, vol. 23, pp. 5048-5079, May 1981.

Numerical methods for Bloch-Poisson type equations

Norbert J. Mauser*, Christian Ringhofer** and Gabriele Ulich*

* Techn. Univ. Berlin, MA 6-2, Str. d. 17. Juni 136, D-10623 Berlin 12

** Dept. of Math., Arizona State Univ., Tempe, AZ 85287

Abstract

We present numerical methods for a rather new model to calculate quantum steady states of electrons in thermal equilibrium. The well known Bloch equation is generalized to arbitrary distribution functions (Fermi-Dirac) and appropriately coupled to the Poisson equation. A rational approximation of the distribution function yields an iterative algorithm.

I Introduction

We present and discuss algorithms to calculate quantum steady states of thermally equilibrated electrons in a semiconductor device. We consider a one-electron approximation with a self-consistent potential describing the Coulomb interaction of the electrons with each other and with the background of (positive) ions.

The most evident model is the Schrödinger-Poisson approach which has been widely used (see e.g. [1], [2], [3], [4], [5]). However, the solution of the eigenvalue problem in an iterative loop is numerically rather expensive, such that it might be interesting to look for alternatives.

It is well known that under the assumption of Boltzmann statistics the steady state density matrix can be obtained from the so called Bloch equation (see e.g. [6], [7], [8]). The Bloch equation is a parabolic equation in which the role of the time is played by the reciprocal of the temperature. The spatial operator is the negative Hamiltonian and the initial datum is a delta-distribution. The charge density is obtained by evaluating the density matrix at the normalized diagonal. The appropriate coupling to the Poisson equation leads to a non-standard evolution problem ([8], [9]).

We present the generalization of the linear Bloch equation to a self-consistent model for arbitrary occupation probabilities. A rational approximation of the distribution function yields an iteration in the temperature variable, the position variable can be discretized e.g. by spectral methods. Due to the specific coupling to the Poisson equation Bloch-type equations give the steady state density matrix only for one, fixed temperature.

II Modeling

In the usual approach the state of the system is described by one-particle wave functions in some L^2 -space (e.g. [1], [2], [3], [4], [5]). One has to calculate the eigenfunctions of the stationary Schrödinger equation

$$(II.1) \quad H\psi_i(r) = E_i\psi_i(r), \quad \|\psi_i\|_{L^2} = 1, \quad i \in N_0$$

The Hamilton operator reads

$$(II.2) \quad H = -\frac{\hbar^2}{2m}\Delta + V(r)$$

with some local potential $V(r)$ taking into account the direct Coulomb interaction of the electrons. Additional terms like an applied external potential, heterojunction discontinuities, exchange-correlation terms etc. can be included without changing the algorithms. The (electron) density is given by

$$(II.3) \quad n(r) = \sum_{i \in N_0} f(E_i - F, \beta) \bar{\psi}_i(r) \psi_i(r)$$

where $f(E_i - F, \beta)$ is the probability of finding an electron in the state ψ_i . Thermodynamics enter via the reciprocal temperature $\beta = 1/kT$, where k is the Boltzmann constant and T the temperature.

In order to achieve self-consistency we calculate $V(r)$ as a function of $n(r)$ by simply solving the Poisson equation. Note that this direct coupling yields a β -dependence of all quantities.

For electrons we should use Fermi - Dirac (FD) statistics :

$$(II.4) \quad f_{FD}(E_i - F, \beta) = \frac{1}{e^{\beta(E_i - F)} + 1}$$

which are approximated by Boltzmann statistics for high temperatures (small β)

$$(II.5) \quad f_B(E_i - F, \beta) = e^{\beta(E_i - F)} = Z e^{-\beta E_i}.$$

The "Fermi energy F " and Z are (β -dependent) normalization constants :

$$(II.6) \quad \int n(x) dx = 1$$

As an alternative to the Schrödinger-Poisson equation the (charge) density $n(x)$ for a given temperature β can be expressed using what we call "Bloch-ansatz" :

$$(II.7) \quad n(x) = \text{Tr } f(H - F \cdot id)$$

Here we mean "trace" in the following sense :

$$(II.8) \quad n(x) = \rho(r = x, s = x, \beta^*) \equiv \rho(x, x)$$

where $\rho(r, s, \beta^*)$ is the (normalised) density matrix, i.e. the integral kernel of the density operator $f(H - F \cdot id, \beta^*)$. We have

$$(II.9) \quad \rho(r, s, \beta^*) = f(H - F \cdot id, \beta^*) \delta(r - s)$$

or

$$(II.10) \quad \rho(r, s, \beta^*) = \sum_{j=1}^{\infty} f(E_j - F, \beta^*) \bar{\psi}_j(r) \psi_j(s)$$

which shows the equivalence of the Bloch-ansatz to the Schrödinger approach.

For the Boltzmann distribution (II.7) yields the "classical" Bloch equation [6] as an evolution equation (in β) for the unnormalized density matrix.

$$(II.11) \quad \frac{\partial}{\partial \beta} \rho(r, s, \beta) = -H_r \rho(r, s, \beta); \quad \beta \in (0, \beta^*)$$

where H_r means acting on the r variable only.

$$(II.12) \quad \rho(r, s, \beta = 0) = \delta(r - s)$$

$$(II.13) \quad n(r) = \frac{\rho(r, r, \beta^*)}{\int \rho(y, y, \beta^*) dy}, \quad \beta^* \text{ fixed}$$

We had to fix β^* since the coupling to the Poisson equation requires attention : A direct coupling for temperature dependent $n(x, \beta) := \rho(x, x, \beta)$ yields a β -dependent Hamiltonian and the Bloch equation is not valid. Only $\rho(r, s, \beta^*)$ has a physical meaning for the selfconsistent Bloch equation ([8], [9]).

For other than the Boltzmann distribution no such parabolic evolution equations can be derived. However, we can still derive iterations for the density matrix without solving the eigenvalue problem (II.1), (II.2).

III Numerical methods

For the Boltzmann statistics case the following relaxed iteration for the potential has been proven to be convergent in [8] : For given V_k the Bloch equation (II.11), (II.12) is solved for $\beta \in (0, \beta^*)$, the density (II.13) is evaluated for fixed $\beta = \beta^*$ and the Poisson equation yields a potential \tilde{V}_{k+1} which gives the new potential

$$(III.14) \quad V_{k+1}(x) = (1 - \omega)V_k(x) + \omega\tilde{V}_{k+1}(x), \quad \omega < 1$$

For Fermi-Dirac statistics it is possible to use an analogous iteration which avoids the explicit calculation of the Fermi-level [10] :

$$(III.15) \quad \rho_k(r, s, \beta)$$

\Downarrow

$$(III.16) \quad n[\rho_k](x) = \rho(x, x, \beta^*)$$

\Downarrow

$$(III.17) \quad \Delta V_k[\rho_k](x) = -n[\rho_k](x)$$

$$H_k[\rho_k] = -\Delta + V_k[\rho_k]$$

\Downarrow

$$(III.18) \quad \frac{\partial}{\partial \beta} \tilde{\rho} = e^{-\beta \frac{1}{2}(H_r + H_s)}, \quad \beta \in (0, \beta^*)$$

$$(III.19) \quad \tilde{\rho}(\beta = 0) = \delta(r - s) - \rho_k(r, s, \beta = 0)$$

\Downarrow

$$(III.20) \quad \rho_{k+1}(r, s, \beta) = \frac{\tilde{\rho}(r, s, \beta)}{\int \tilde{\rho}(y, y, \beta) dy}$$

Like in the Boltzmann case the iteration has to be appropriately underrelaxed to assure convergence. Underrelaxed iterations have been used for the Schrödinger -Poisson equation ([4], [5]).

However, the numerical results of Nier [4] show that the optimal underrelaxation parameter ω depends heavily on the Debye length λ_D of the considered device and that ω has to be extremely small for realistic values of λ_D . On the other hand, the iteration (III.15) - (III.20) applies only for F-D statistics as given by (II.4), which looks different for the 2-D case (see e.g. [1]).

The method we propose is based on a rational (Padé)-approximation of an arbitrary distribution function $f(\beta(E_i - F))$:

$$(III.21) \quad f(z) \simeq \tilde{f}(z) = \prod_{\ell=1}^N \frac{a_{\ell}z + b_{\ell}}{z + c_{\ell}}$$

To understand the idea we take the Boltzmann case $f(\cdot) = e^{-\beta z}$ and solve the Bloch equation (II.11), (II.12) with an implicit Euler-scheme :

$$(III.22) \quad (I + \Delta\beta H_r)\rho_{\ell+1} = \rho_{\ell}, \quad k = 0, \dots, N-1, \quad N \cdot \Delta\beta = \beta^*$$

$$(III.23) \quad \rho_0 = \delta(r - s)$$

Hence we see that the numerical iteration corresponds to the approximation

$$(III.24) \quad f(z) = e^{-\beta^* z} \simeq (1 + \frac{\beta^*}{N}z)^{-N} = \tilde{f}(z) = \prod_{\ell=1}^N \frac{\frac{N}{\beta^*}}{z + \frac{N}{\beta^*}}$$

In general we have

$$(III.25) \quad \tilde{f}(\beta^*(H - F)) = \prod_{\ell=1}^N (H - F \cdot id + c_{\ell} \cdot id)^{-1} (a_{\ell}H - a_{\ell}F \cdot id + b_{\ell})$$

where β^* is contained in the coefficients a_{ℓ} , b_{ℓ} , c_{ℓ} which have to be determined once from the (Padé) - approximation.

Hence we use the following iteration for the potential V and the Fermi-level F :

$$(III.26) \quad V_k(x), F_k \quad (F_{k-1})$$

\Downarrow

$$H_k = -\Delta + V_k$$

$$(III.27) \quad (a_{\ell}H_k - a_{\ell}F_k \cdot id + b_{\ell})(H_k - F_k \cdot id + c_{\ell} \cdot id)^{-1} z_{\ell+1} = z_{\ell}, \quad \ell = 0, \dots, N-1$$

$$(III.28) \quad z_0 = \delta(r - s)$$

$$(III.29) \quad \rho_k[V_k, F_k](r, s, \beta^*) := z_{N-1}(r, s)$$

\Downarrow

$$(III.30) \quad n_k[V_k, F_k](x) = \rho_k(x, x, \beta^*)$$

\Downarrow

$$(III.31) \quad C_k[V_k, F_k] := \int n_k[V_k, F_k](x) dx$$

\Downarrow

$$(III.32) \quad V_{k+1}(x) \text{ solving } -\Delta V_{k+1}(x) = \frac{n_k[V_k, F_k](x)}{C_k[V_k, F_k]}$$

\Downarrow

$$(III.33) \quad F_{k+1} \text{ is given implicitly by the normalization (II.6)}$$

A secant scheme using C_k and also F_{k-1} , C_{k-1} gives :

$$(III.34) \quad F_{k+1} = F_k - (C_k - 1) \frac{F_k - F_{k-1}}{C_k - C_{k-1}}$$

Note that in general the iteration (III.26)–(III.33) does not correspond to any differential equation like the Bloch equation for Boltzmann statistics.

The above iteration has been tested for a 1-D model. The discretization in x has been done both with finite differences on an equidistant grid and with a spectral collocation method using trigonometric basic functions. Homogeneous Dirichlet boundary conditions have been used.

First results have shown a good convergence with a moderate number (below 10) of iterates. Indeed, the Padé-approximation with $N = 6$ is basically indistinguishable from the exact FD-distribution. The implementation of a 2-D code using the spectral method is ongoing.

References

- [1] C. M. Snowden *Simulation of microwave and millimeter-wave semiconductor devices* Proc. IWCE 92, Beckman Inst., pp. 13–18, 1992
- [2] T. Kerkhoven, M. Raschke and U. Ravaioli, *Self-consistent simulation of corrugated layered structures* Superlatt. and Microstruct., Vol. 12, No. 4, pp. 505–508, 1992
- [3] H. Arman and R. Khoie *A self-consistent multisubband model for calculation of scattering rates in quantum well structures* Proc. IWCE 92, pp. 175–179, 1992
- [4] F. Nier *Étude Numérique d'Algorithmes pour la Résolution du Système Schrödinger Poisson*, Ph.D. Thesis of Ecole Polytechnique, pp. 107–129, 1991
- [5] B. Vinter *Subbands and charge control in a two-dimensional electron gas field effect transistor* Appl. Phys. Lett., Vol. 44, p. 307, 1984
- [6] R. P. Feynman *Statistical mechanics: a set of lectures* W. A. Benjamin (ed.) Frontiers in Physics, London 1972
- [7] W. Lui and J. Frey *A simplified method for quantum size effect analysis in submicron devices* J. Appl. Phys., Vol. 64 (12) pp. 6790–6794, 1988
- [8] A. Arnold, P. A. Markowich and N. Mauser *The one-dimensional periodic Bloch-Poisson equation* Math. Mod. and Meth. in Appl. Sciences, Vol 1 pp. 83–112, 1991
- [9] N.J. Mauser and C. Ringhofer *Quantum Steady States via self-consistent Bloch-type equations* Proc. IWCE 92, Beckman Inst., pp. 273–276, 1992
- [10] A. Arnold *Private Communication*, 1992

Transport Simulations for Quantum Well Heterostructures and Lasers

Matt Grupen and Karl Hess

Beckman Institute for Advanced Science and Technology,
University of Illinois,
Urbana, IL 61801

Abstract

A review of the simulation of electronic transport at and over quantum well heterostructures is given. A large number of experimental results relating to thermionic emission, phonon-bottlenecks, electron-electron interaction, tunneling, as well as spontaneous and stimulated emission will be reviewed in the presentation, while this short note will concentrate on effects pertinent to quantum well lasers.

I. Introduction

Transport over a quantum well and recombination of electrons (holes) in the well form, in principle, a formidable problem. Obviously the transport in the immediate vicinity of the well can only be treated by quantum methods, while in the remainder of the structure (laser) the transport is well described in a semiclassical way using drift-diffusion theory and extended versions of it.[1, 2] To bring classical and quantum transport regions together requires sophisticated theories such as the Bardeen Transfer Hamiltonian approach[3] or the Landauer-Büttiker theory[4, 5] extended to include inelastic processes. Under ordinary circumstances a rate equation approach and rates such as those used by Brum and Bastard[6] may be sufficient. However, subtleties with the Pauli Principle and the ranges and normalizations of classical and quantum regions present problems.[7]

In the presentation related to this paper, a number of cases of generation-recombination and transport near quantum wells is going to be discussed. In this paper, however, we discuss only a few facts relevant to the development of complete tools for semiconductor laser simulation. It is in the modulation response of lasers that the subtleties of quantum well transport most obviously manifest themselves. The model we propose is not to be viewed as a final version but as a starting point for a complete numerical approach.

II. The Multiscale Problem

As mentioned, transport in heterostructures such as a quantum well laser poses a difficult problem. The quantum well must be treated appropriately because it is the region of positive gain, and, therefore, it is the most critical determinate of optical output. However, the regions surrounding the quantum well also play important roles in the device characteristics. For example, the current blocking regions significantly influence the effective quantum efficiency of the device. Also, the transport in the separate confinement regions can lead to low frequency roll-off in the modulation response[8] and gain saturation.[9, 10]

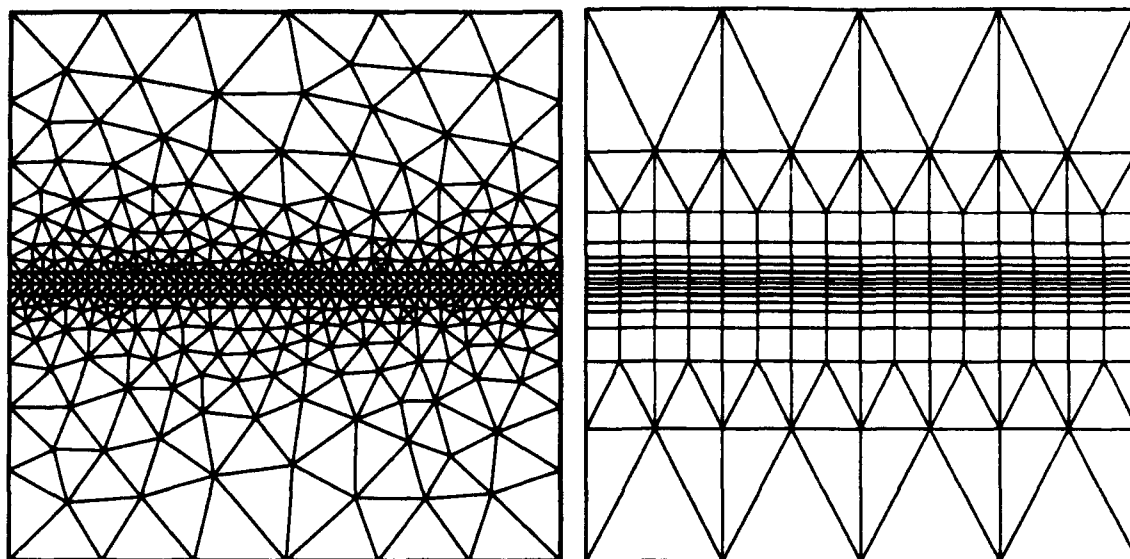


Figure 1: Triangular tessellation of a quantum well device is inappropriate due to a small aspect ratio. A combination of rectangles and triangles produces a mesh consistent with the quasi-one-dimensional nature of quantum wells.

As a result, the simulation of such structures is also a multiscale problem.

The first difficulty with such a multiscale problem involves the tessellation of the device. It is standard practice to use triangles to discretize a complicated device structure because they conform well to nonrectangular features, such as contoured current blocking regions. However, in the case of quantum well structures, triangles have a serious drawback. A well formed triangle usually has an aspect ratio on the order of one. If a ridge waveguide laser structure with a $10\text{ }\mu\text{m}$ wide and $100\text{ }\text{\AA}$ thick quantum well is being simulated, then the use of triangles leads to an explosion in the number of mesh points. Rectangles, on the other hand, lend themselves very well to the quasi-one-dimensional nature of quantum wells. As a result, the use of rectangles and triangles produces the most convenient tessellation for a quantum well device. Figure 1 shows two meshes generated for a simple quantum well structure. The more flexible aspect ratio of the rectangles produces a much more appropriate mesh.

III. Coupling the Classical and Quantum Regions

In addition to optimizing the number of mesh points, the use of rectangles in and near the quantum well is crucial to the treatment of transport in these regions. Rectangles allow the quantum well to be divided up into transverse cross-sections. The mesh points in each cross-section can then be used to solve a one-dimensional Schrödinger's Equation, giving the eigenenergies and wavefunctions for the bound states in the well. Transport in and around the well can then be treated in a special way. Figure 2 is a schematic representation

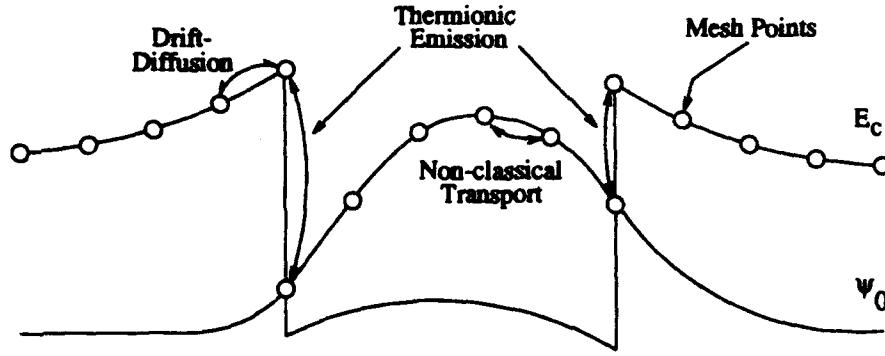


Figure 2: The schematic diagram shows a transverse cross-section of the quantum well. The classical and quantum mechanisms that determine the carrier distributions are labelled where they apply. Note that the circles represent mesh points in the discretization, not charge carriers.

of this treatment. The figure shows the conduction band edge and the ground state wavefunction for one of the transverse cross-sections of the well. Drift-diffusion theory describes the transport up to the quantum well. The particle fluxes are expressed in the following form.

$$\begin{aligned}\vec{j}_n &= -\mu_n N_c \mathcal{F}_0(\eta_n) (T \nabla \eta_n + \nabla E_c) \\ \vec{j}_p &= -\mu_p N_v \mathcal{F}_0(\eta_p) (T \nabla \eta_p - \nabla E_v)\end{aligned}$$

where,

$$\eta_n = (F_n - E_c)/T, \eta_p = (E_v - F_p)/T$$

$$\mathcal{F}_j = \frac{1}{j!} \int_0^\infty dx \frac{x^j}{1 + e^{x-\eta}}$$

In each of the flux equations, the first term in parentheses is the diffusion term and the second is the drift term, where the gradient in the band edge determines the field. Thermionic emission theory then determines the flux of carriers into or out of the well according to the following expression.

$$j = A_2^* T^2 \frac{2m_1 m_2}{m_1 + m_2} \left\{ \exp\left(\frac{E_{Fn2} - E_{C2}}{kT}\right) - \exp\left(\frac{E_{Fn1} - E_{C1} - E_B}{kT}\right) \right\}$$

where, A_2^* = the Richardson constant.

Inside the quantum well, classical transport is not valid in the transverse direction; it is the form of the wavefunctions that determines the distribution of carriers inside the well. When carriers are injected into the well, they conform to the distribution established by the wavefunctions almost instantaneously. Classical transport is eliminated by requiring that the quasi-Fermi levels inside the well are constant in the transverse direction. Their absolute positions are determined by the carrier fluxes into or out of the well. Consequently, it is the self-consistent agreement between drift-diffusion and thermionic

emission transport that ultimately determines the filling of the well and, thus, the gain of the laser. Inclusion of quantum reflections and resonances represents a straight forward extension of this treatment.

Classical transport is, however, permitted in the lateral direction. And again, the rectangular tessellation enables this to be treated correctly. The lateral transport is treated with drift-diffusion theory. The divergence in the carrier flux is calculated for each point in a transverse cross-section of the well, but these divergences are added together to determine the continuity equation for the wavefunction as a whole. The result is the propagation along the quantum well of entire wavefunctions that represent the carrier distribution, and maintaining the shape of the wavefunctions as carriers propagate is necessary for the correct simulation of quantum systems.

The importance of coupling the classical and quantum regions correctly is evident in figure 3. This figure shows the modulation responses for two different lasers. Each laser has 5000 Å *AlGaAs* separate confinement regions on each side of a 100 Å *GaAs* quantum well. However, for the device in the top figure, the optical confining region was ungraded, whereas for the bottom device, it was linearly graded. The ungraded device shows low frequency roll-off and increased saturation of the resonant peaks when compared to the graded device. The roll-off is due to slow carrier drift in the separate confinement region, and the gain saturation is the result of diffusive diode current. Even though the active regions are identical in the two devices, the optical output characteristics are very different due to transport in the surrounding bulk regions. Consequently, it is critical that the classical regions are properly coupled to the quantum region in order to accurately calculate the device performance.

IV. Accounting for Carrier Capture

Although the method described above is an effective way of coupling the classical regions of a laser with the quantized active region, there is an element missing that may affect the measured characteristics of the device. We discussed the way in which thermionic emission theory is used to determine the filling of the well. Thermionic emission theory assumes that the carriers on one side of a heterojunction are in thermal equilibrium with themselves but not with carriers on the other side of the junction. This is valid, but it also assumes that once a carrier passes to the other side of the junction it immediately relaxes and is in thermal equilibrium with carriers on that side. It does not consider the finite time it takes for the hot injected carrier to lose its excess energy. This time is typically on the order of a picosecond or less. However, due to the self-consistent agreement of the thermionic emission current injected into the well and the drift-diffusion current in the surrounding regions, any finite hot carrier concentration in the well can lead to increased carrier densities in the separate confinement regions.[7] This increase in carrier density will produce a diffusive barrier to transport and increase gain saturation, thereby affecting the laser characteristics.

The coupling of the classical and quantum regions described above can be extended to account for the relaxation of hot quantum carriers into bound states. Figure 4 schematically represents this extension. As before, the conduction band edge and the ground

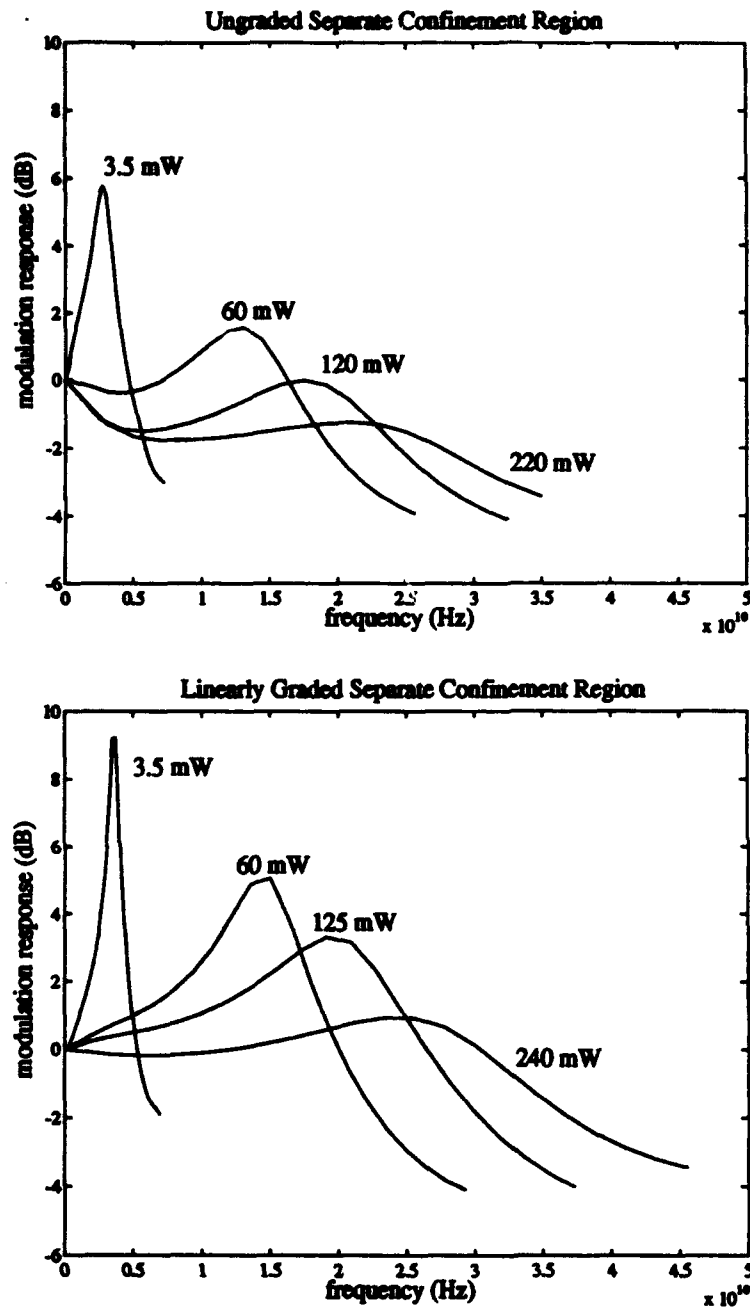


Figure 3: The modulation responses for two different 100 Å *GaAs* quantum well lasers. Each device has 5000 Å *AlGaAs* separate confinement regions on each side of the quantum well. However, the optical confinement region for the top device is ungraded while the confinement region for the bottom device is linearly graded.

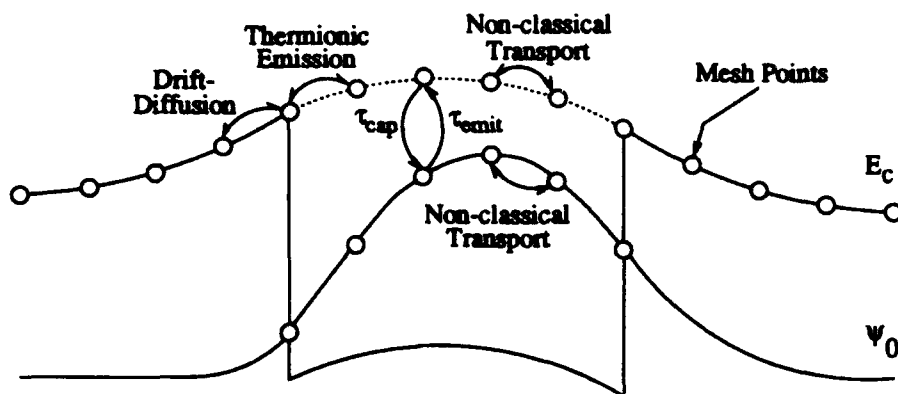


Figure 4: The schematic diagram shows a transverse cross-section of the quantum well. The classical and quantum mechanisms that determine the carrier distributions are labelled along with the scattering between quasi-bound and bound quantum states.

state wavefunction for a transverse cross-section of the quantum well is shown. Like the previous method, drift-diffusion theory is used to determine carrier transport up to the well, and thermionic emission theory is used to determine injection into the well region. However, now carriers are not injected into bound states but rather into higher energy, quasi-bound states. This is achieved by simply using thermionic emission with zero barrier height. The quasi-bound states are not assumed to be in thermal equilibrium with the bound states in the quantum well. They have a separate quasi-Fermi level to determine their occupancy. The quasi-bound states exchange carriers with the bound states through a scattering lifetime, e.g. as calculated in [6]. The result will be a finite hot quantum carrier concentration, the consequence of which can only be determined by the self-consistent solution of the transport equation throughout the rest of the device.

V. Conclusion

The characteristics of quantum well devices are most critically dependent on the properties of the well, which can be accurately treated only with quantum mechanical methods. However, the device performance can also depend strongly on transport in the remainder of the device, as we have shown with the modulation response of semiconductor lasers. Transport in these other regions is most tractable when treated with classical theories, but this leads to the problem of coupling the quantum treatment of the active region with the classical treatment of the surrounding regions. To do this correctly requires complex mesoscopic theories which are not easily implemented in computer simulation. In this paper, we have presented a model which serves as a starting point for the numerical solution of this difficult problem.

Acknowledgement

This work was supported by the National Science Foundation through the ERC and

through NCCE at the University of Illinois, Urbana-Champaign.

References

- [1] G.H. Song, K. Hess, T. Kerkhoven, and U. Ravaioli, "Two-Dimensional Simulation of Quantum Well Lasers," *European Trans. on Telecom.*, Vol. 1, No. 4, pp. 375-381, July-Aug. 1990.
- [2] M. Grupen and K. Hess, "Simulation of Transport over Heterojunctions," *Simulation of Semiconductor Devices and Processes*, Vol. 4, pp. 303-311, Sept. 1991.
- [3] J. Bardeen, "Tunnelling from a Many-Particle Point of View," *Phys. Rev. Lett.*, Vol. 6, No. 2, pp. 57-59, Jan. 1961.
- [4] R. Landauer, "Spatial Variation of Currents and Fields Due to Localized Scatterers in Metallic Conduction," *IBM J. Res. Develop.*, Vol. 1, pp. 223-231, July, 1957.
- [5] M. Büttiker, "Four-Terminal Phase-Coherent Conductance," *Phys. Rev. Lett.*, Vol. 57, No. 14, pp. 1761-1764, Oct. 1986.
- [6] J.A. Brum and G. Bastard, "Resonant Carrier Capture by Semiconductor Quantum Wells," *Phys. Rev. B*, Vol. 33, No. 2, pp. 1420-1423, Jan. 1986.
- [7] S.C. Kan, D. Vassilovski, T.C. Wu, and K.Y. Lau, "Quantum Capture Limited Modulation Bandwidth of Quantum Well, Wire, and Dot Lasers," *Appl. Phys. Lett.*, Vol. 62, No. 19, pp. 2307-2309, May 1993.
- [8] R. Nagarajan, T. Fukushima, M. Ishikawa, J.E. Bowers, R.S. Geels, and L.A. Coldren, "Transport Limits in High-Speed Quantum-Well Lasers: Experiment and Theory," *IEEE Phot. Tech. Lett.*, Vol. 4, No. 2, pp. 121-123, Feb. 1992.
- [9] G.W. Taylor and P.R. Claisse, "Effects of Drift and Diffusion Current Flow on the High-Speed Performance of Quantum Well Lasers," *Appl. Phys. Lett.*, Vol. 62, No. 7, pp. 723-725, Feb. 1993.
- [10] M. Grupen and K. Hess, "The Self-Consistent Simulation of the Modulation Responses of Quantum Well Lasers," 51st Annual Device Research Conference, Session IIB, June 21, 1993.

Quantum Simulation of Several-Particle Systems

P. Douglas Tougaw, Craig S. Lent, and Wolfgang Porod

Department of Electrical Engineering

University of Notre Dame

Notre Dame, IN 46556 U.S.A.

Abstract

We study the simulation of quantum cellular automata and how such a simulation is simplified by the features of FORTRAN-90. We demonstrate the use of second-quantized operators to write the cell Hamiltonian and explain the Hartree self-consistent method for simulating a many-cell system. Finally, several examples of simulated QCA devices are shown.

1. Introduction

The particular quantum system we simulate consists of several of the quantum cells shown in Fig. (1a). We determine the ground state of the system by solving the time-independent Schrödinger equation. Each cell consists of five coupled quantum dots which contain a total of two electrons. The cells only interact with each other Coulombically; no tunneling is allowed between cells. Since the state of each cell is affected by its nearest neighbors, we call such a system a quantum cellular automaton (QCA).

Because of Coulombic repulsion between the two electrons in each cell, the charge density exhibits strongly bistable behavior. The ground state of the cell is therefore in one of the two states shown in Fig. (1b). Because of this bistable nature, we can use the state of each cell to encode binary information. We define a cell polarization which measures to what extent the cell is in one of the two stable states shown in Fig. (1b):

$$P \equiv \frac{(\rho_1 + \rho_3) - (\rho_2 + \rho_4)}{\rho_0 + \rho_1 + \rho_2 + \rho_3 + \rho_4} \quad (1)$$

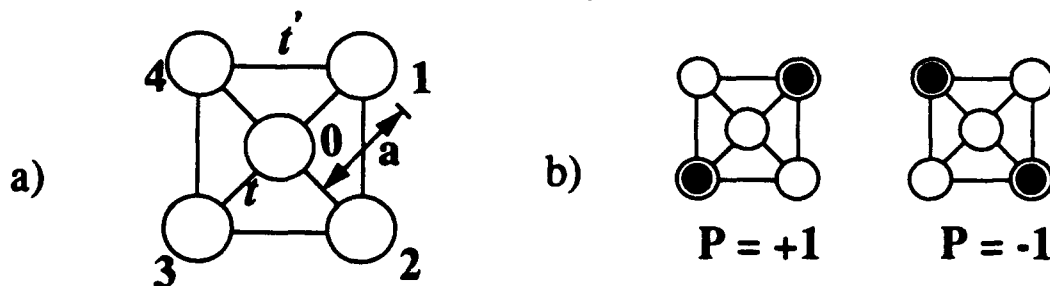


Figure 1. Schematic of the basic five-site cell. (a) The geometry of the cell with $t=0.3\text{meV}$, $t'=t/10$, and $a=20\text{ nm}$. (b) Coulombic repulsion causes the electrons to occupy antipodal sites within the cell. The two bistable states have cell polarizations of $P=+1$ and $P=-1$ (See equation (1)).

II. Second-Quantized Hamiltonian

We define a second-quantized annihilation operator, $\hat{a}_{i,\sigma}$, which destroys an electron on site i with spin σ , and a creation operator, $\hat{a}_{i,\sigma}^\dagger$, which creates a particle on site i with spin σ . The product of these two operators, $n_{i,\sigma} = \hat{a}_{i,\sigma}^\dagger \hat{a}_{i,\sigma}$, is the number operator for that site and spin.

Using these operators we can compactly write the Hubbard-type tight-binding Hamiltonian of a single isolated cell:

$$H_0^{cell} = \sum_{i,\sigma} E_0 n_{i,\sigma} + \sum_{i>j,\sigma} t_{i,j} (a_{i,\sigma}^\dagger a_{j,\sigma} + a_{j,\sigma}^\dagger a_{i,\sigma}) + \sum_i E_Q n_{i,\sigma} n_{i,\sigma'} + \sum_{i>j,\sigma,\sigma'} V_Q \frac{n_{i,\sigma} n_{j,\sigma'}}{|\vec{R}_i - \vec{R}_j|} \quad (2)$$

This Hamiltonian includes on-site energies, tunneling between sites, on-site charging costs, and Coulombic repulsion between each pair of sites. The interaction with neighboring cells alters the on-site energies in the first term.

III. Unique Features of Fortran-90

Fortran-90 supports a level of data abstraction sufficient to allow direct implementation of these second-quantized operators and the related state vectors. We have created user-defined types representing creation and annihilation operators and many-electron site kets and bras. We also provide functions to convert between these types and to define the effect of each operator on all other data types.

The second useful feature of Fortran-90 for our purpose is operator overloading. This allows us to use an operator without regard to the data types it acts upon. We then provide an interface that invokes the appropriate function based on the data types involved. In this case, the action of a creation or annihilation operator on a Dirac ket in the site representation is specified. The operation is "overloaded" onto the normal multiplication symbol. A similar overloading specifies that multiplication of a Dirac bra and ket be interpreted as the inner product of the state vector.

Fig. (2) shows a segment of Fortran-90 code that demonstrates how easily quantum mechanical expressions can be written as Fortran code once these definitions are in place. The code is easily understandable because the level of data abstraction matches the level of quantum mechanical abstraction.

$$\langle i_1 | \sum_{i,j,\sigma} (a_{i,\sigma}^\dagger a_{j,\sigma} + a_{j,\sigma}^\dagger a_{i,\sigma}) | i_2 \rangle$$

↓

```
! -- Hopping terms --
do ist=1,NSTATES
  do jst=1,NSTATES
    do ispl=1,2
      H(i1,i2)=H(i1,i2)+ket2bra(baseket(i1))*(T0(ist)* &
        &ad(ist,ispl)*a(jst,ispl)*baseket(i2))
      H(i1,i2)=H(i1,i2)+ket2bra(baseket(i1))*(T0(ist)* &
        &ad(jst,ispl)*a(ist,ispl)*baseket(i2))
    end do
  end do
end do
```

Figure 2. Conversion from second-quantized quantum mechanical expression to FORTRAN-90 code using data abstraction techniques.

IV. Hartree Self-Consistent Calculations

Since electrons are not allowed to tunnel between cells, we can solve for the ground state of each cell separately. Such intracellular calculation includes exchange and correlation effects exactly. The interaction between cells is included using a Hartree self-consistent technique. Once the iterative solution of the system has converged, the system is in an eigenstate. Use of several different initial conditions and comparison of the eigen-energies allows us to determine the overall ground state of the system.

V. Application: Quantum Cellular Automata

We have used this scheme to simulate many arrangements of quantum cells. The most fundamental of these calculations is shown in Fig. (3a). The system consists of two cells as shown in the inset. The charge density of cell 2 is fixed, and the induced polarization in cell 1 is then calculated. This is repeated for many values of P_2 in the range $[-1,+1]$ and the induced polarization P_1 can then be calculated as a function of P_2 . This cell-cell response function demonstrates the highly nonlinear and bistable nature of the cell's response to its neighbors.

Fig. (3b) shows a similar cell-cell response function calculated at several non-zero temperatures. This requires calculating the thermal expectation value of the polarization over the canonical ensemble. As seen in the figure, the nonlinearity of the response degrades as the temperature increases.

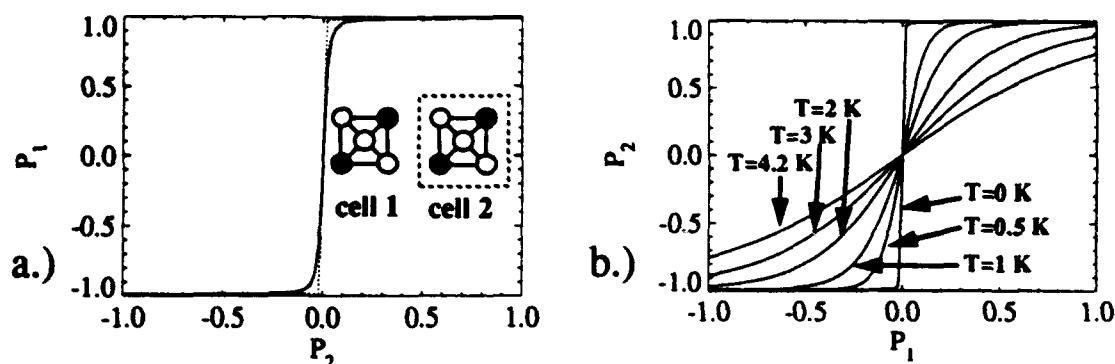


Figure 3. Cell-cell response functions. (a) The cell-cell response function for the basic five-site cell at zero temperature. The solid line corresponds to the antisymmetric case, and the dotted line to the symmetric case. (b) The temperature dependence of the cell-cell response function.

Fig. (4) shows three of the basic QCA devices. Fig. (4a) demonstrates that even a weakly polarized cell can drive a line of similar cells and that the bistable saturation of the cell-cell response will return the signal to maximum polarization in subsequent cells. Fig. (4b) demonstrates that a signal will propagate through a right-angle turn without degradation, and Fig. (4c) shows that a single line of cells can fan out to multiple lines and maintain signal integrity.

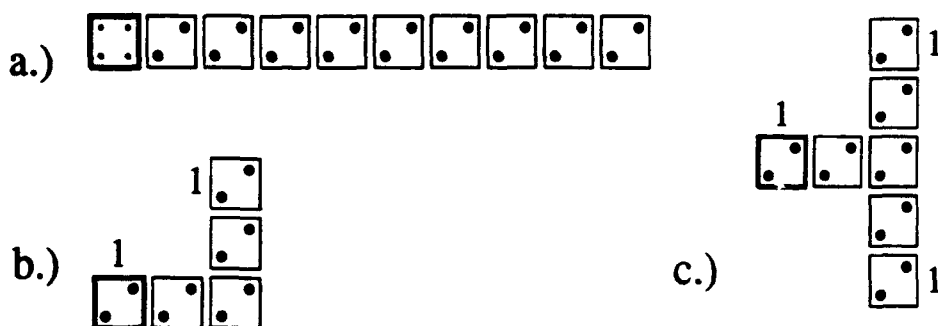


Figure 4. Three basic QCA devices. (a) A line of cells can be used to transmit information from one point to another. (b) The signal is transmitted correctly around a corner. (c) A single line can fan out correctly to multiple lines with the same signal. These are not schematic diagrams; they are plots of the actual results of the self-consistent calculation of the ground state of each system. The diameter of each dot is proportional to the charge density on the site. The cells with heavy borders are fixed; all others are free to react to the fixed charge.

Fig. (5) shows that the state of a free cell matches the majority of its fixed neighbors. This majority voting logic can provide the basis of a new computing architecture. If one of the fixed neighbors is called the "program line", such an arrangement of cells can be interpreted to be a programmable AND/OR gate. The program line determines the nature of the gate (AND vs. OR), and the other inputs are applied to the gate thus defined.

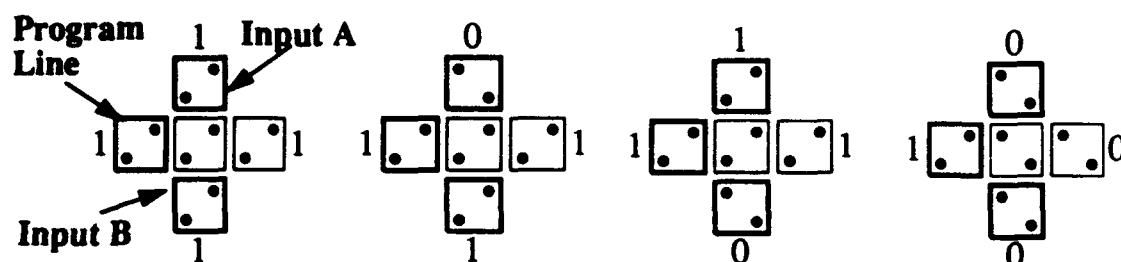


Figure 5. The programmable AND/OR gate. Here, the program line is set to one, so the other two inputs are being applied to an OR gate. In each of the four cases, the output is one if either of the two inputs are one. This is a plot of the result of a self-consistent ground state calculation.

Finally, Fig. (6) shows how to cross two lines of cells without having the signals interfere. Wire crossings are very important for implementation of devices like adders and exclusive-OR gates. Such a quasi-two-dimensional crossing is impossible with conventional devices.

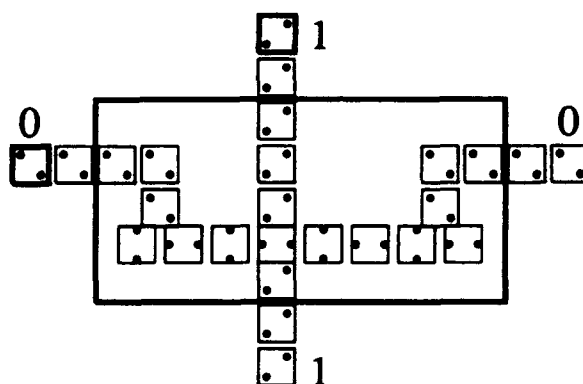


Figure 6. One way to cross two QCA wires without signal interference. The box shows the extent of the crossing, so a system designer can simply place such an arrangement of cells wherever two wires need to be crossed. This is the result of a self-consistent ground state calculation for this system.

Acknowledgment

This work was supported in part by AFOSR, ARPA, ONR, and a National Science Foundation Graduate Fellowship.

References

- [1] C. S. Lent, P. Douglas Tougaw, and Wolfgang Porod, *Appl. Phys. Lett.*, **62**, 714 (1993).
- [2] C. S. Lent, P. Douglas Tougaw, Wolfgang Porod and Gary H. Bernstein, *Nanotechnology* **4**, 49 (1993).
- [3] P. Douglas Tougaw and C.S. Lent, to appear in *Journal of Applied Physics*, **74** (1993).
- [4] C. S. Lent, P. Douglas Tougaw, and Wolfgang Porod, submitted to *Journal of Applied Physics* (unpublished).

Full Scale Self Consistent Simulation of Quantum Devices

Dejan Jovanovic and Jean-Pierre Leburton

Beckman Institute for Advanced Science and Technology,
University of Illinois,
Urbana, IL 61801

Abstract

A comprehensive self-consistent Schrödinger-Poisson simulation is presented. The model takes into account the various regions of reduced dimensionality throughout a particular structure which can consist of 0D, 1D, 2D electron gases or bulk. The feasibility of this level of simulation is made possible by an iterative extraction orthogonalization method for solving the Schrödinger equation. This method is superior to conventional eigenvalue solvers since it generates an arbitrary number of eigenstates and easily couples to the Poisson equation. High order effects such as exchange/correlation and single-electron charging are also included in the model. Transport is treated in the linear response regime and used to investigate the Coulomb blockade oscillations observed by Meirav[1].

I. Introduction

Recent advances in semiconductor fabrication techniques have resulted in a large variety of new experimental devices exhibiting quantum effects due to high degrees of confinement. In order to provide a realistic model for the physical analysis and design of these structures, we have developed a comprehensive self-consistent simulation tool that merges the statistical and quantum mechanical aspects of the problem. In this paper, we give the theoretical background of the model and its application to the analysis of the single-electron charging effects on transport through a quantum dot[1]. Previous efforts at simulating quantum structures have focused on localized regions in the device which exhibit reduced degrees of dimensionality. Although these simulations have yielded much information on the general quantum-mechanical properties in a device active region[2, 3, 4], their failure to integrate the dots (0D), leads (1D), contacts (2D), and bulk regions in a particular geometry prevents them from achieving good quantitative agreement with experimental data. This lack of coherence has been due in large part to computational time constraints in solving the Schrödinger equation over an arbitrarily large number of grid-points, N_G . Conventional eigenvalue solvers typically scale as N_G^3 and generate N_G eigenvectors thereby restricting their application to problems with small grids. In a general quantum device, however, eigenvalue problems need to be solved in all regions exhibiting reduced dimensionality and require a large N_G . In addition, only a few eigenenergies are often necessary owing to the small number of occupied eigenstates in a typical device. Clearly, a robust eigenvalue solver that addresses the above issues and allows easy self-consistent coupling to the Poisson equation is needed to bridge the gap between "localized" modeling and full scale device simulations.

II. Theoretical Background

We solve the Schrödinger equation with the iterative extraction orthogonalization method (IEOM) that propagates a set of eigenstates according to [5]

$$|\phi_n^i\rangle = e^{-\alpha \hat{H}} |\phi_n^{i-1}\rangle, \quad i^{\text{th}} \text{ iteration} \quad (1)$$

where \hat{H} is the Hamiltonian and n indicates a particular eigenstate. The parameter α is selected so as to minimize the error in the Taylor expansion of the exponential operator $\exp(-\alpha \hat{H})$. In general, N_i iterations are required to maintain the accuracy of the Taylor expansion and eliminate all the error projections of the basis states $|m\rangle$ of \hat{H} onto the initial guess state $|\phi_n^0\rangle$. Gram-Schmidt orthonormalization is performed over the entire set of states after each iteration to prevent any excited states from collapsing to the ground state. This procedure also eliminates all projections of states with $m < n$ such that after N_i iterations,

$$|\phi_n^{N_i}\rangle \simeq |n\rangle + \sum_{m=n+1} |m\rangle e^{-\alpha N_i (E_m - E_n)} \frac{\langle m | \phi_n^0 \rangle}{\langle n | \phi_n^0 \rangle}. \quad (2)$$

The repeated exponential scalings of the error projections and Gram-Schmidt orthonormalizations therefore eventually convert $|\phi_n^0\rangle$ into a pure basis state $|n\rangle$. The convergence criterion for the algorithm is determined by the expression

$$\alpha N_i = \max \left[\frac{1}{E_m - E_n} \log \left(\frac{\langle m | \phi_n^0 \rangle}{\epsilon \langle n | \phi_n^0 \rangle} \right) \right] \quad (3)$$

where ϵ is an imposed error tolerance. Clearly, α should be chosen to be as large as possible to reduce the number of iterations and still allow a high degree of accuracy in the Taylor expansion of $\exp(-\alpha \hat{H})$. Eq. 3 points out an overall limitation of this method in that the number of iterations required to achieve convergence scales inversely with the energy separation between eigenstates. This problem can be alleviated somewhat by selecting initial states with the appropriate symmetry such that the error projections vanish for eigenstates with opposite parity. Following previous time dependent treatments [6, 7, 8, 9], the exponential operator in Eq. 1 is typically cast into a split form which separates its potential and kinetic energy components. The execution time of each iteration depends on how efficiently the discretized kinetic energy portion can be solved. For the rectilinear geometries often encountered in quantum devices, the kinetic energy operator can be further separated into (x, y, z) components which can be treated independently and solved by rapid elliptical solvers. In general, each application of the propagator scales with N_G and the Gram-Schmidt algorithm scales with $N_G N_E^2$ where N_E is the number of eigenenergies required in a particular simulation. If N_E is relatively small, as is often the case in nanostructures, $N_G N_E^2 \ll N_G^3$ and the IEOM shows a significant improvement over conventional eigenvalue solvers. The chief advantage offered by the rapid convergence of this method is the ability to accurately treat the dimensionality of each region in a quantum device out to an appropriate set of boundaries. We solve the 3D Schrödinger equation in the quantum dot (0D) regions, the 2D Schrödinger equation in the lead (1D) regions, and 1D solutions are obtained in the contact (2D) regions. Semiclassical solutions are

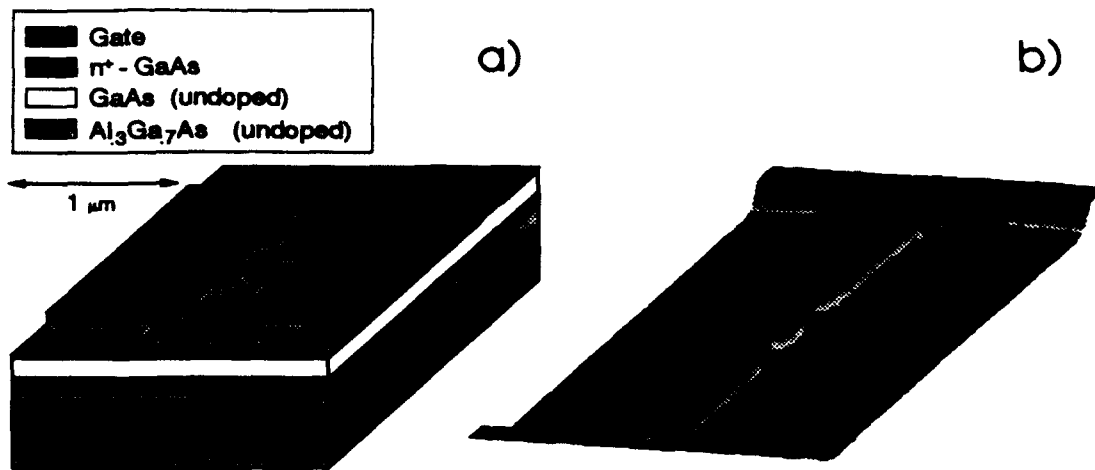


Figure 1: (a) Experimental device geometry used for investigating single electron charging effects. A negative bias on the gate confines charge in the lateral direction to form 1D leads and a quantum dot between the restrictions. Modulation of the bottom gate bias adds electrons to the quantum dot one at a time. (b) Theoretical electron density for a slice taken at the device active region. The formation of a quantum dot is visible as an island of charge between the two 1D leads.

used for carriers that do not exhibit confinement (holes) and regions that are not governed by quantum mechanics. Charge densities are constructed by scaling the amplitude squared of each wavefunction with its appropriate statistical weight[10]. The distribution functions used in the scaling of the 0D localized states are derived from the grand canonical ensemble with the constraint that only an integer number N^* of electrons occupy the quantum dot. N^* is determined for a given gate bias by free energy minimization[11]. The potential $\phi(\vec{r})$ in each region is inherently coupled by the Poisson equation which maintains continuity of $\phi(\vec{r})$ and its first derivative. The Schrödinger and Poisson equations are solved self-consistently by a modified Newton method that incorporates a line-search step to allow convergence at low temperatures. In addition, exchange and correlation effects are self-consistently treated with the Kohn-Sham approach[12] using Perdew-Zunger parameterization for the correlation potential[13]. Presently, the simulation is carried out in equilibrium although in principle, nonequilibrium solutions are possible for accurate evaluations of the quantum-mechanical current. We therefore compute the quantum transport properties of devices in the linear response regime which is typically the case for Coulomb blockade measurements.

III. Results

We have applied the simulation to the analysis of single electron charging effects in the inverted semiconductor-insulator-semiconductor structure reported by Meirav et al[1]. Fig. 1 shows the specific device geometry used in our investigation along with a charge density surface taken at the device active region. A negative bias on the top gate confines charge

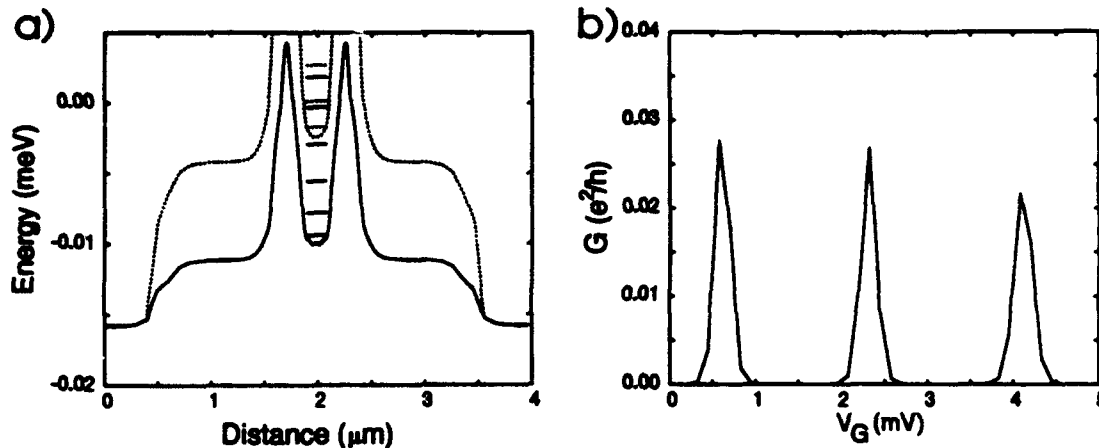


Figure 2: (a) The quasi-1D eigenenergies and their respective localized 0D eigenenergies. The eigenenergies associated with the first (solid) and second (dotted) quasi-1D states are the only ones occupied ($E_F = 0 \text{ meV}$). The formation of sharp barriers leads to tunneling into and out of the quantum dot. (b) The transport characteristics of the device exhibit sharp thermally broadened ($T = 50 \text{ mK}$, $L_o = .8 \mu\text{m}$) peaks that correspond to the addition of a single electron to the dot and the subsequent lifting of the Coulomb blockade.

in the lateral direction to form 1D leads and a quantum dot between the restrictions. The quantum dot is visible in Fig. 1b as an island of charge between the two 1D leads. In addition, the presence of charge in the leads and contacts is also apparent. As the bias on the backgate is increased, electrons are added to the dot in single increments. This results from the electrostatic Coulomb repulsion between localized electrons which modifies the occupation probability of each level and allows only an integer number of electrons to occupy the dot. The 0D region is delineated by sharp barriers separated by a distance L_o which bring the dot into weak contact with the leads via tunneling. Fig. 2a shows the localized eigenenergies in the dot superimposed over their respective quasi-1D adiabatic eigenenergies obtained by extending the simulation of the leads into the quantum dot region. Although two quasi-1D modes have access to the dot, the upper channel is essentially closed owing to its relatively wide tunnel barriers and low statistical weight. The device should therefore exhibit single mode characteristics. Transport in the device is calculated using the appropriate expression derived for the Coulomb blockade regime[14, 15]. Transmission probabilities are evaluated using a transfer matrix calculation based on the quasi-1D eigenenergy as a function of distance. The charge imbalance caused by single electron occupation is modeled by minimizing the free energy with respect to the number of electrons. The difference between successive free energies then gives the charging energy required to add a single electron to the dot for that gate bias[14]. Transport characteristics of the device (Fig. 2) exhibit sharp thermally broadened peaks that correspond to the addition of a single electron to the dot and the subsequent lifting of the Coulomb blockade. The peak amplitude and periodicity show good agreement with the experimental data[1].

IV. Conclusion

In summary, we have demonstrated the viability of a comprehensive device-scale simulation tool for analyzing quantum devices. The lengthy eigenvalue calculation, which is the chief obstacle to this level of simulation via conventional methods, has been overcome by employing the rapidly converging iterative extraction orthogonalization method that can treat an arbitrary number of eigenenergies and wavefunctions. Finally, transport characteristics for a device exhibiting single-electron charging effects were obtained and shown to exhibit good agreement with experimental findings.

Acknowledgements

The authors would like to thank Tetsuo Kawamura, Paul von Allmen, Jeffery Bigelow, Massimo Macucci, and Karl Hess for their helpful advice. Support for this work was provided by NSF Grant ECS 91-20641 and the Joint Services Electronics Program.

References

- [1] U. Meirav, M. A. Kastner, and S. J. Wind, *Phys. Rev. Lett.* **65**, 771 (1990).
U. Meirav, P.L. McEuen, M.A. Kastner, E.B. Foxman, A. Kumar, and S.J. Wind, *Z. Phys. B* **85**, 357 (1991).
- [2] S. E. Laux and F. Stern, *Appl. Phys. Lett.* **49**, 91 (1986).
- [3] A. Kumar, S. E. Laux, and F. Stern, *Phys. Rev. B* **42**, 5166 (1990).
- [4] D. Jovanovic and J. P. Leburton, *IEEE Electron Device Lett.* **14**, 7 (1993).
- [5] R. Kosloff and H. Tal-Ezer, *Chem. Phys. Lett.* **127**, 223 (1986).
- [6] M. D. Feit, J. A. Fleck, Jr., and A. Steiger, *J. Comp. Phys.* **47**, 412 (1982).
- [7] R. Kosloff, *J. Phys. Chem.* **92**, 2087 (1988).
- [8] M. H. Degani, *Appl. Phys. Lett.* **59**, 57 (1991).
- [9] J. Bigelow, J. P. Leburton, and M. Degani, *Proc. Int. Workshop Computational Electron.* Urbana, IL., 309 (1992).
- [10] D. Jovanovic and J. P. Leburton, to be published
- [11] H. Van Houten, C.W.J. Beenakker, and A. A. M. Staring, in *Single Charge Tunneling, NATO Advanced Study Institute, Series B: Physics*, edited by H. Grabert and M.H. Devoret (Plenum, New York, 1991).
- [12] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- [13] J. P. Perdew and A. Zunger, *Phys. Rev. B* **23**, 5048 (1981).
- [14] C. W. J. Beenakker, *Phys. Rev. B* **44**, 1646 (1991).
- [15] Y. Meir, N. S. Wingreen, and P. A. Lee, *Phys. Rev. Lett.* **66**, 3048 (1991).

Planar Supercell Simulations of 3D Quantum Transport in Semiconductor Nanostructures

D. Z.-Y. Ting, S. K. Kirby, and T. C. McGill

Thomas J. Watson, Sr., Laboratory of Applied Physics
California Institute of Technology
Pasadena, California 91125, U.S.A.

Abstract

We introduce the planar supercell method as a means for treating 3D quantum transport in mesoscopic tunnel structures. The flexibility of the method allows us to examine a variety of physical phenomena relevant to quantum transport, including alloy disorder, interface roughness, defect impurities, and 0D, 1D, and 2D quantum confinement, in device geometries ranging from double barrier heterostructures to quantum wire electron waveguides. As examples, we examined quantum transport in double barrier structures with interface roughness, and in quantum wire electron waveguides.

I. Introduction

In modeling quantum transport in semiconductor resonant tunneling heterostructures, one can often assume perfect periodicity in the lateral directions, thereby reducing the mathematical description to a 1D problem in which only the potential variation along the growth direction need be considered. However, in realistic device structures we need to take into account imperfections such as interface roughness, impurities, and alloy disorder which are incompatible with the assumption of translational invariance in the parallel directions. In this work, we introduce the planar supercell method as a general purpose model for treating these structural imperfections. The model is designed for flexibility so that it can be used not only to study tunnel structures such as the double barrier heterostructure, but also lower dimensional mesoscopic devices such as quantum wire electron waveguides. Using this method, we have studied how the transport properties of double barrier heterostructures are influenced by interface roughness. We also briefly report on a study of transport in quantum wire electron waveguides.

II. Method

We use a planar supercell tight-binding Hamiltonian and specify a structure as a stack of N_z layers perpendicular to the z -direction, with each layer containing a periodic array of rectangular planar supercells of $N_x \times N_y$ sites. Within each planar supercell, the potential assumes lateral variations as dictated by the device geometry. Figure 1 illustrates a set of planar supercell stacks used in simulating double barrier structures with interface roughness. Our model is formally equivalent to the one-band effective mass equation[1]

$$-\frac{\hbar^2}{2} \nabla \cdot \frac{1}{m(\mathbf{x})} \nabla \psi + V(\mathbf{x})\psi = E\psi, \quad (1)$$

discretized over a Cartesian grid, and subject to periodic boundary conditions (with su-

percell periodicity) in the x - and y -directions, and open boundary conditions in the z -direction. The transmission coefficients for structures described by the planar supercell stack can be determined by the direct application of the multiband method developed by Ting *et al.* [2]

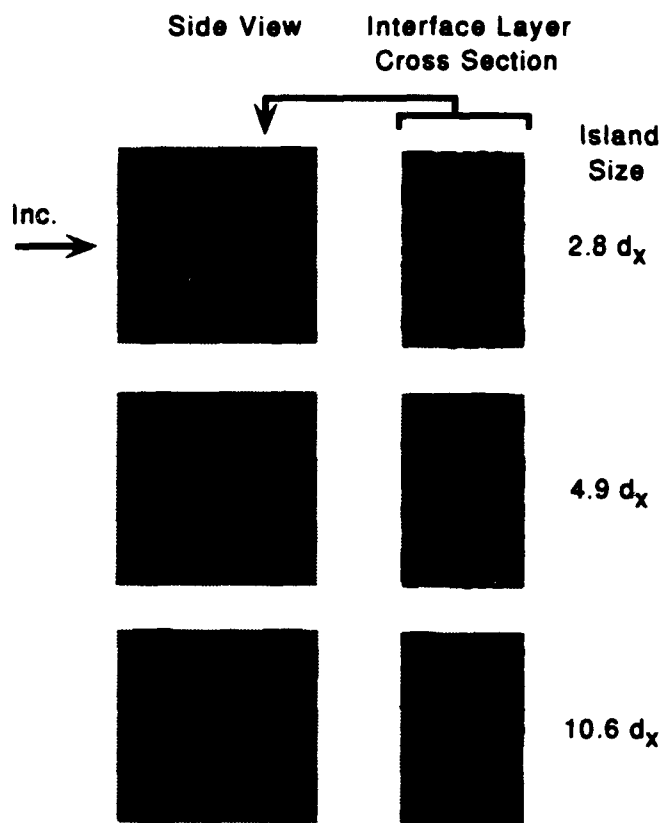


Fig. 1. A set of planar supercell stacks used in simulating double barrier structures with interface roughness. Light and dark areas represent sites occupied by well and barrier materials, respectively. Cross sections shown are the 25×16 planar supercells representing the rough interfacial layers between the quantum well and the second barrier. Note that the supercells are repeated in the lateral directions.

III. Applications

To illustrate the applications of the planar supercell method, we examine the effect of interface roughness on transmission properties of double barrier structures, and effects of neutr. impurities and geometric imperfections in quantum wire electron waveguides.

Interface Roughness

Figure 2 shows transmission spectra near the $n = 1$ resonance for a set of GaAs/AlAs double barriers structures with interface roughness as depicted in Figure 1. For each GaAs-AlAs interface on the incident side, a 50% random coverage rough interfacial layer is placed between the pure GaAs layers and the pure AlAs layers. Random configurations of interfacial layers with different island sizes are generated with a simulated annealing

algorithm.[3] Three of the structures have average island sizes (λ) of 28 Å, 49 Å, and 106 Å; an additional reference structure with smooth virtual crystal approximation (VCA) $\text{Al}_{0.5}\text{Ga}_{0.5}\text{As}$ alloy interfacial layers is included for comparison. We note that the spectra for structures with rough interfaces show a series of satellite peaks not present in the reference structure spectrum. The satellite peak strength increases with island size, but peak positions are approximately the same for all three rough structures. The satellite peaks are the result of interface roughness induced k_{\parallel} scattering of off-resonance states into on-resonance states. In principle, interface roughness can scatter a given k_{\parallel} into a continuous range of k'_{\parallel} . But due to the finite supercell size, k_{\parallel} can scatter only into a discrete set of $N_z \times N_y$ parallel k vectors differing from it by a reciprocal lattice vector. Therefore the contributions from the scattered states appear as individual satellite peaks. The spacing between satellite peaks can be decreased with larger supercell sizes. For sufficiently large supercells, the spacings would become smaller than the resonance peak width, and the satellite peaks would then coalesce.

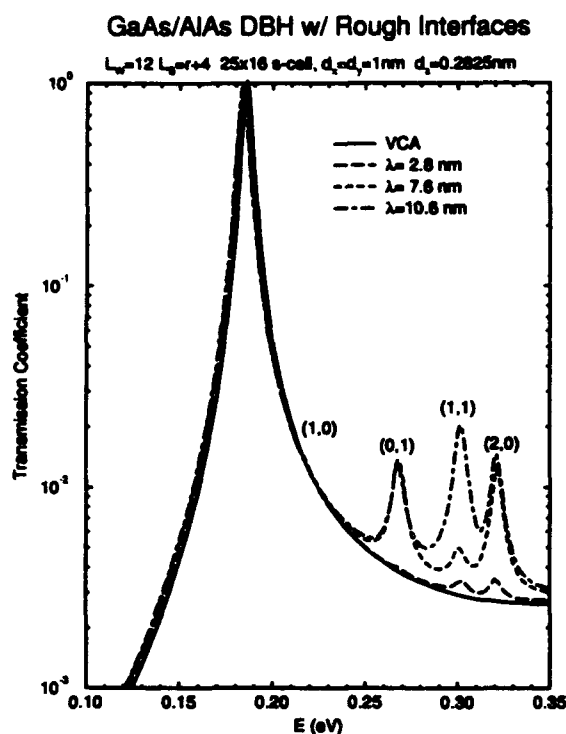


Fig. 2. Transmission coefficients for a set of GaAs/AlAs double barrier structures with rough interfaces as depicted in Fig. 1.

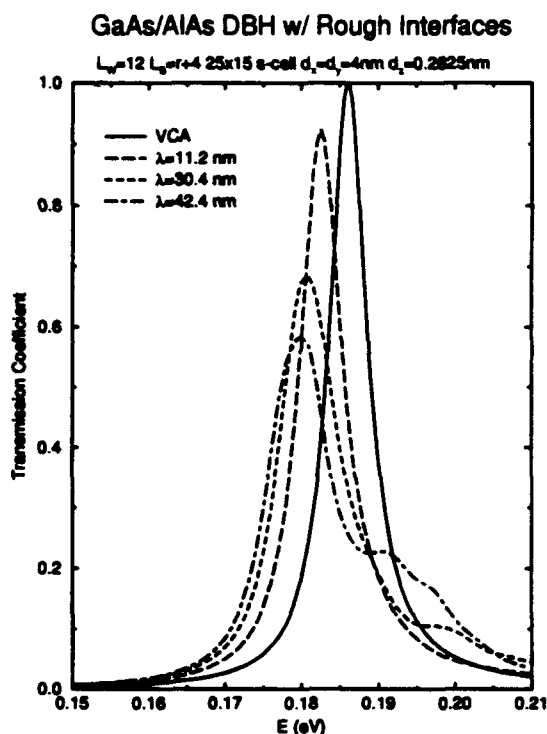


Fig. 3. Transmission coefficients for structures similar to those shown in Fig. 2, except that the island sizes are larger.

The scattering mechanism described above accounts for how interface roughness can reduce peak-to-valley ratios in current-voltage characteristics of double barrier resonant tunneling diodes by increasing off-resonance transmission coefficients. The other major effect of interface roughness is the broadening of resonance peaks through wave function localization. In Figure 3 we show transmission spectra for a set of structures similar to those in the previous example, but with larger island sizes of 112 Å, 304 Å, and 424 Å.

We note that the $n = 1$ transmission resonance down-shifts and broadens with increasing island size. This can be explained in terms of wave function localization. The presence of the rough interfacial layer between the quantum well and the second barrier introduces well width fluctuation, and in effect divides the structure into wide-well ($L_W = 13$, in this example) and narrow-well ($L_W = 12$) regions. If the island sizes are sufficiently large (compared to the de Broglie wavelength), it is possible to have quantum well states whose wave functions are laterally localized in the wide-well regions. Since the wide-well regions are also the narrow-barrier regions (second barrier width of $L_B = 4$ rather than 5), the transmission resonance associated with these localized states should be down-shifted (due to the wider well), and broadened (due to the narrower barrier).

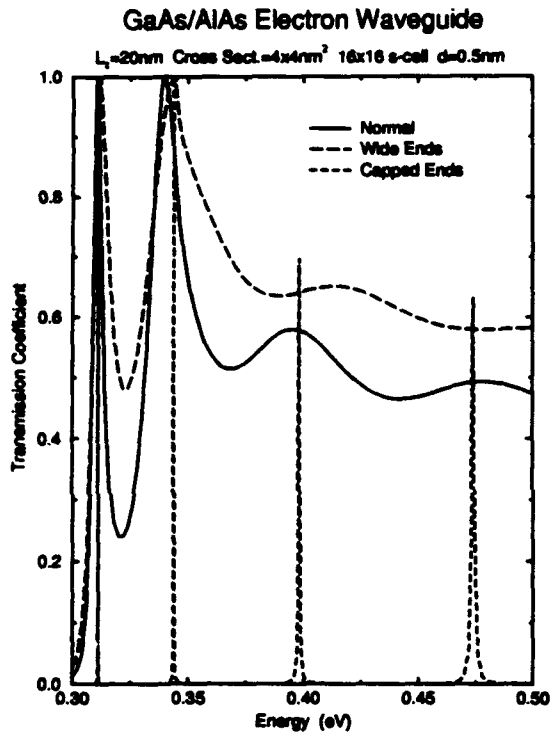


Fig. 4. Transmission coefficients for a set of quantum wire electron waveguide with different waveguide opening geometries.

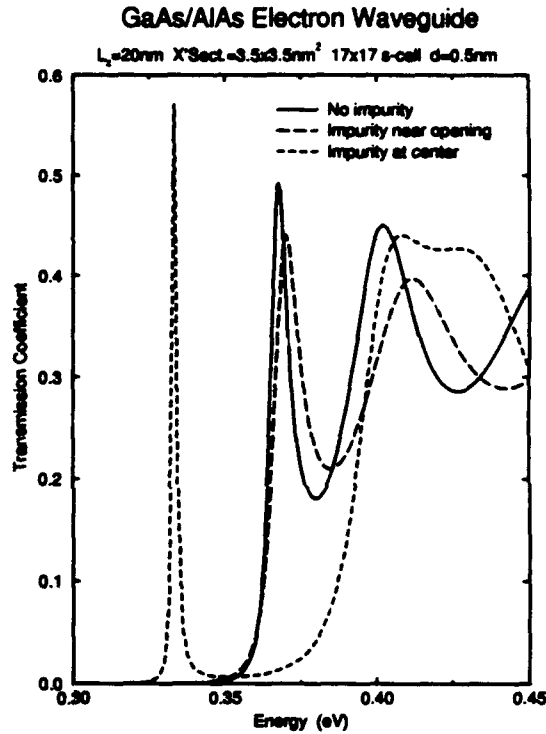


Fig. 5. Transmission coefficients a set of quantum wire electron waveguide with and without neutral impurity.

Electron Waveguide

We explore the transmission properties of quantum wire electron waveguides with the planar supercell method. Using discretization sizes of $d_x = d_y = d_z = 5\text{\AA}$ and 16×16 planar supercells, we modeled a GaAs quantum wire structure 200\AA in length, $40\text{\AA} \times 40\text{\AA}$ in cross section, surrounded on the sides by AlAs walls, and the ends by GaAs electrodes. To study the sensitivity of transport properties to waveguide opening geometry, we have modeled two similar structures. The first has slightly wider waveguide openings : at the two ends of the waveguide, the cross section is widened to $60\text{\AA} \times 60\text{\AA}$ for the first 5\AA (lengthwise), and $50\text{\AA} \times 50\text{\AA}$ for the next 5\AA ; the rest of the wire remains $40\text{\AA} \times 40\text{\AA}$ in cross section. The other structure is obtained by simply capping the ends of the original

structure with 15 Å thick AlAs layers. Figure 4 shows the transmission spectra for the three structures described above. We see that the amplitude of the transmission coefficients and the resonance positions and widths vary considerably in these structures.

In the final example we study the effect of neutral impurities on transport properties of electron waveguides. We model three waveguide structures : one with no impurities, another with an attractive impurity at the center of the waveguide, and a third with an impurity near the waveguide opening. Figure 5 provides a picture of how an impurity perturbs the waveguide transmission spectrum. We note that an impurity near the waveguide opening does not strongly perturb the lowest two resonances since these modes have small probability densities near the waveguide opening. However, an impurity at the waveguide center down-shifts and narrows the lowest ($n=1$) resonance while leaving the $n=2$ mode relatively unperturbed. (Note that the $n=3$ mode has also been down-shifted to near the $n=2$ resonance.) This is because the $n=1$ resonance has the largest probability density at the center of the waveguide, while the $n=2$ mode has a node at the center.

IV. Summary

We discussed the planar supercell method as a means for treating 3D quantum transport in mesoscopic tunnel structures. We demonstrated its applications with studies of interface roughness in double barrier heterostructures, and geometric imperfections in electron waveguides. In addition to the examples discussed here, we have also use our method to study the effect of alloy disorder and impurity scattering in various device geometries. We found the flexibility of the method to be extremely useful in exploring a wide range of issues relevant to the operation of tunnel devices.

Acknowledgment

The authors would like to thank Eric Van de Velde, W. L. Johnson, and W. R. Frensley for helpful discussions. SKK would like to thank the Office of Naval Research for graduate fellowship support during this research. This work was supported by the U. S. Office of Naval Research (ONR) under Grant No. N00014-89-J-1141.

References

- [1] D. J. BenDaniel and C. B. Duke, Phys. Rev. **152**, 683 (1966).
- [2] D. Z.-Y. Ting, E. T. Yu, and T. C. McGill, Phys. Rev. B **45**, 3583 (1992).
- [3] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

Simulation of coherent quantum transport in a magnetic field: recovery of conductance quantization

Manhua Leng and Craig S. Lent

Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556
USA

Abstract

We extend the Quantum Transmitting Boundary Method (QTBM), a numerical algorithm for solving the two-dimensional Schrödinger equation for scattering states, to include an applied magnetic field to 2D systems. We apply this technique to simulate the magneto-transport of a periodically corrugated electron channel. The conductance quantization of such structure is recovered when the channel is long. The index of this quantization is a non-monotonic function of energy.

I. Introduction

For structures of ultra-submicron dimension, electron transport is in the quantum regime. The shapes of these devices and the potential landscapes can be flexibly tailored, for example, by electro-static gating on top of the 2DEGs. Understanding the details of transport in these structures inevitably involves numerical solutions of the 2D Schrödinger equation.

Figure 1 schematically shows the type of structures we consider: the region of interest can be partitioned into several input and output lead regions ($\Omega_i, \Omega_j, \dots$) and a "device region" (Ω_0), where scattering states are to be solved. A perpendicular magnetic field is present. For device modeling, a mode matching approach is widely adopted [1,2]. The problem with this method is that, though useful for the most simple structures, it is difficult to apply for arbitrary potential profiles. A general numerical algorithm, the Quantum Transmitting Bound-

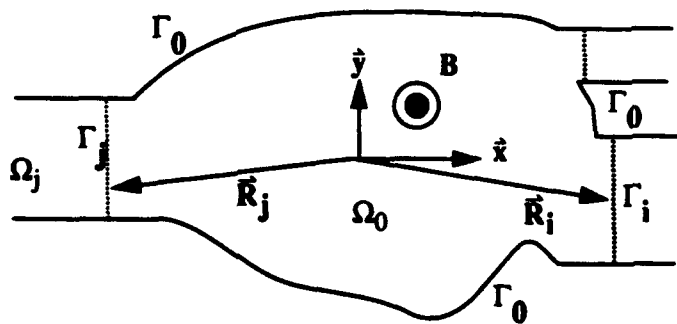


FIGURE 1. The problem geometry. Γ_i is the boundary between the device region Ω_0 and lead i . Γ_0 denotes all the other boundaries, where the wavefunctions vanish. The applied magnetic field B is perpendicular to the 2D system.

ary Method (QTBM) developed by Lent and Kirkner [3], allows more complicated potential profile and can cope with complicated devices. It has proven to be an efficient technique in device modeling in [4].

II. Model And Method

We extend the QTBM to include the effects of applied magnetic fields. With the choice of the vector potential in Landau gauge, $\vec{A} = -By\hat{x}$, the single-band 2D effective-mass Schrödinger equation becomes,

$$\left\{ \frac{-\hbar^2}{2m^*} \nabla^2 + \frac{ie\hbar By}{m^*} + \frac{e^2 B^2 y^2}{2m^*} + V_0(x, y) \right\} \Psi(x, y) = E\Psi(x, y). \quad (1)$$

In the QTBM algorithm, we expand the scattering state in a lead region as a superposition of the local transverse modes, including both traveling and evanescent ones. The modes in the lead regions are obtained by solving (Eq. 1) as a quadratic eigenvalue problem for k at given energy E , using the form $\Psi(x, y) = e^{ikx}\chi(y)$. Part of the difficulty of the problem at non-zero field is that the transverse mode eigenfunction, the $\chi(y)$'s, are not orthogonal, in contrast to the zero field case. Except for the incident modes, the complex amplitude of each mode is an unknown for which we must solve. In the Finite Element Method scheme, we are able to properly implement the boundary conditions and obtain the numerical solutions of full wavefunctions in the whole device by the solution of a single linear problem. We calculate other interested physical quantities of transport, for instance, current density distribution and transmission coefficients directly from the wavefunction. The extended QTBM technique gives us the capability of solving the magneto-transport problem for arbitrarily shaped devices with arbitrary potentials, as that of Figure 1.

III. Application and Results

As an important example, we present the results of this technique applied to a corrugated quantum channel (Figure 2) which illustrates the recovery of conductance quantization for long periodically modulated channels. The conducting width is periodically modulated be-

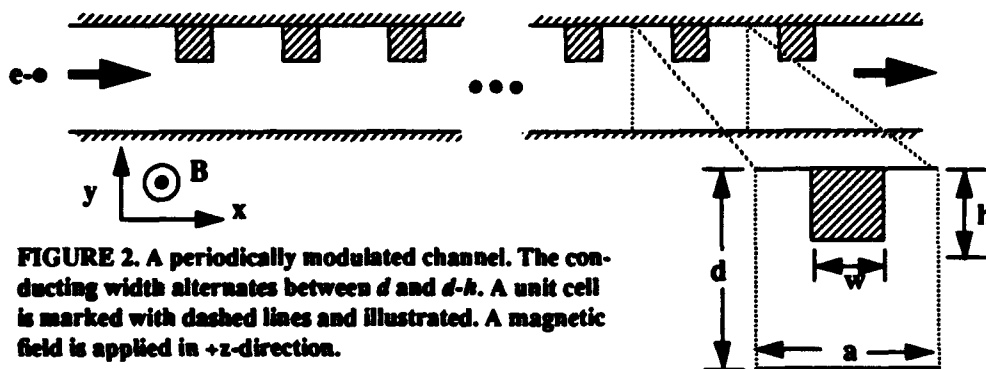


FIGURE 2. A periodically modulated channel. The conducting width alternates between d and $d-h$. A unit cell is marked with dashed lines and illustrated. A magnetic field is applied in $+z$ -direction.

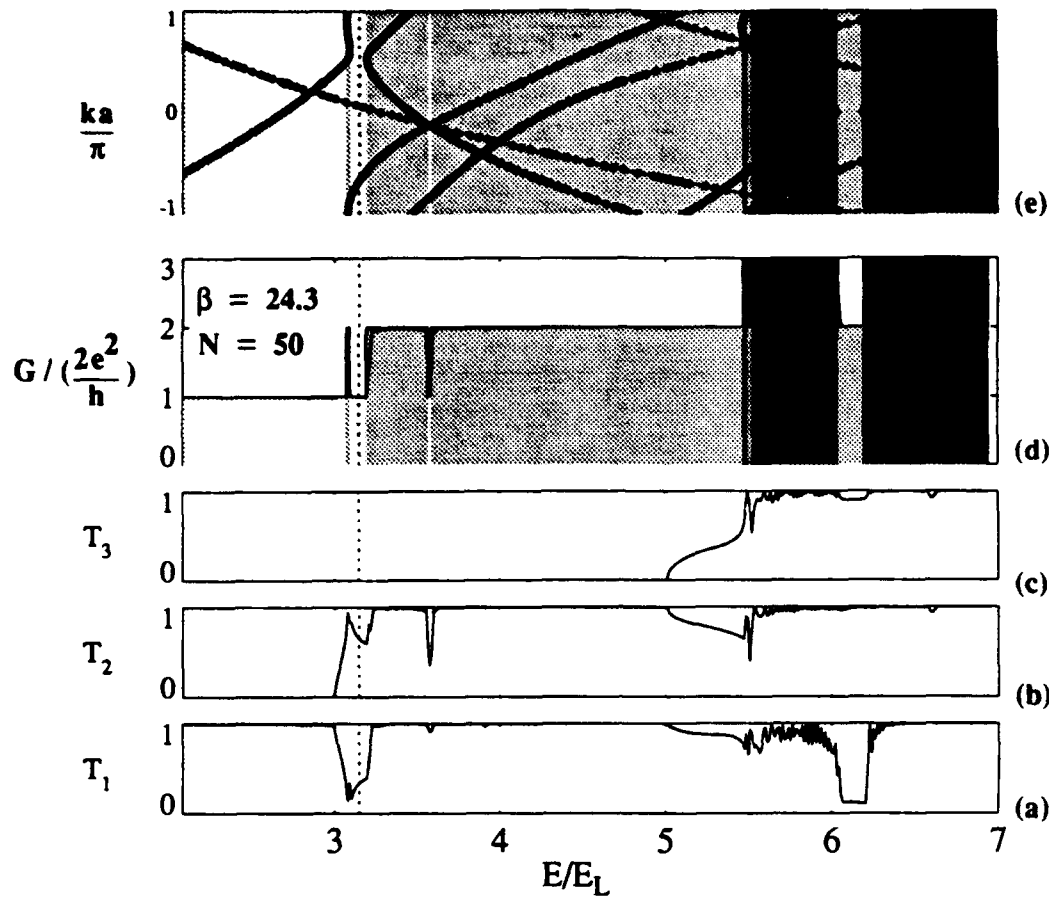


FIGURE 3. Numerical results for the structure in Figure 2 at $d/a=2.0$, $h/a=0.6$, $w/a=0.4$ and $\beta=24.3$. (a) Transmission coefficient for a channel of $N=50$ unit cells. The incident wave is in the *first* traveling mode; (b) the incident is in the *second* mode; (c) the incident is in the *third* mode. (d) The Landauer Conductance, showing the quantization in a non-monotonic way as the function of energy. (e) The bandstructure of an infinite modulated channel. The *index* of the conductance plateau in (d) has an one-to-one correspondence to the *number* of positive-velocity states at a given energy in the band diagram.

tween d and $d-h$ with a periodicity of a and N periods. The full solution of wavefunctions and scattering matrix is obtained for one unit cell (one corrugation) and the scattering matrix for the structure of N unit cells (periods) is calculated using a cascading method. The Landauer two-terminal conductance is then calculated straightforwardly.

Plotted in Figure 3 are the numerical results of transmission and conductance as a function of energy at a modestly high magnetic field. The energy is expressed in units of first bulk Landau level, $E_L = (\hbar\omega_c)/2$, where $\omega_c = (eB)/m^*$ is the cyclotron frequency; the magnetic field is expressed in dimensionless form by $\beta = (da)/l_H^2$, where $l_H^2 = \hbar/(eB)$ is the magnetic length. For the results presented here, the dimensions of the device are set as $d/a = 2.0$, $h/a = 0.6$, $w/a = 0.4$. The effective mass is chosen to be $m^*/m = 0.067$, which is appropriate for GaAs. Hard wall potentials outside the channel and zero inside are chosen for simplicity.

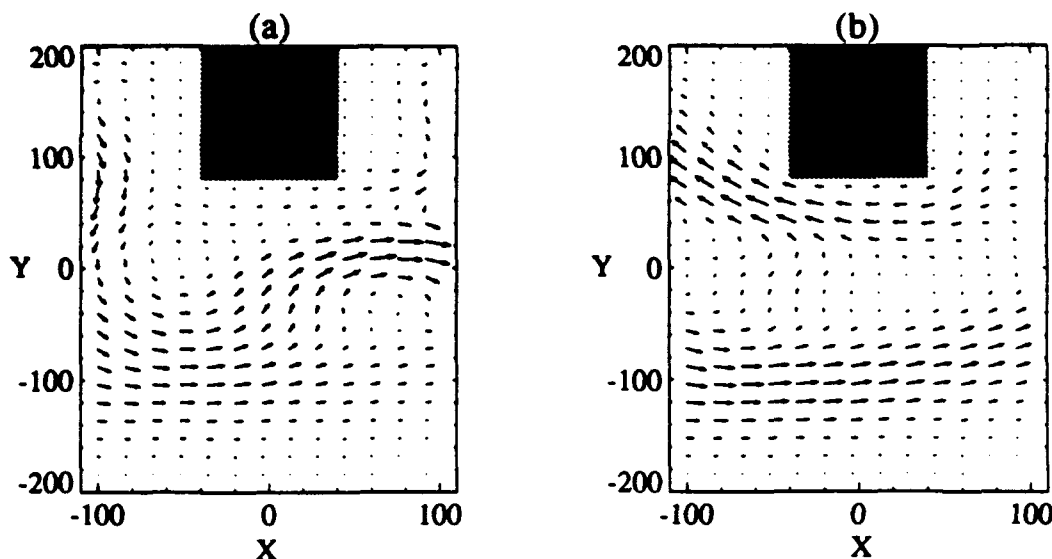


FIGURE 4. Particle current density distribution of the scattering states in the device region of a single unit-cell channel at $E/E_L=3.15$. Highlighted is the potential barrier "finger" which causes the mixing of edge channels in the device. (a) The input is in the first transverse mode; (b) the input is in the second transverse mode.

Drawn in Figure 3(a), (b) and (c) are the mode transmission coefficients, T_1, T_2, T_3 , when the input is in the first, second, and third transverse mode respectively. The channel has 50 periods of modulation. We see a complicated pattern of mode transmission as a function of energy and none of the single mode transmission coefficients is quantized.

In Figure 3(d) we plot the two-terminal Landauer conductance, $G = \frac{2e^2}{h} \sum_i T_i$. It shows that for the long channel of 50 periods, although none of the individual mode transmissions shows the quantization, the total conductance as function of energy is essentially quantized. The height of the plateaus is a non-monotonic function of energy.

In comparison, we also calculated the bandstructure of the infinite periodic system and plot it in the reduced Brillouin zone (Figure 3(e)). We find the index of each conductance plateau has a one-to-one correspondence to the number of positive-velocity (slope) bands in the bandstructure of the infinite problem [5]. The shadings in Figure 3(d) and 3(e) illustrate this correspondence. Full understanding of the variations in conductance from the exact integers is for further study. One reason is that the channel, although long, is still finite.

For an incident energy of $E/E_L=3.15$, marked with the dashed line in Figure 3(a), there are two allowed traveling modes in the lead region. Figures 4 illustrate the particle current density distributions of a one-cell channel. The modulation causes the mixing of edge states in the device region and makes the single mode transmitting pattern complicated. Shown in Figure 4(a) is the particle current density for electrons incident in the first mode (edge-state). The opposite-going edge states on opposite side walls of the channel are weakly coupled in one unit cell, so the resulting transmission coefficient is $T_1=0.96$. Figure 4(b) shows the current for electrons incident in the second mode. The transmission coefficient is $T_2=0.87$.

for one unit cell. For the long channel of 50 unit cells, we have $T_1 = 0.34$ and $T_2 = 0.66$ which sum to unity for the total transmission and make the plateau index of the quantized conductance 1.

VI. Summary

We have extended the QTBM algorithm to include an applied magnetic field to 2D systems for solving current carrying states governed by effective 2D Schrödinger equation. In this algorithm, we have acquired the capability of modeling of magneto-transport with complicated potential profiles and arbitrarily shaped devices. This technique is applied to studying of a periodically corrugated quantum channel. The results illustrate that the total conductance quantization recovers when the modulated channel is sufficiently long. The plateau index is a non-monotonic function of energy. However, the individual mode transmissions are not quantized.

Acknowledgement

This work was supported by the Air Force Office of Scientific Research, by Texas Instruments, and by the National Science Foundation under grant number ECS890025 through the National Center for Computational Electronics, and utilized the Cray -2 at the National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign.

References

- [1] H. Tamura and T. Ando, Phys. Rev. B 44, 1792 (1991).
- [2] R. L. Schult, H. W. Wyld, and D. G. Ravenhall, Phys. Rev. B 41, 12760 (1990).
- [3] Craig S. Lent and David J. Kirkner, J. Appl. Phys. 67, 6353(1990).
- [4] William R. Frensley, Rev. Modern Phys. 62, 745(1990).
- [5] Manhua Leng and Craig S. Lent, submitted to Phys. Rev. Lett.

Numerical Method to Obtain the Two-Dimensional Electronic States for Open Boundary Scattering Problems

Henry K. Harbury and Wolfgang Porod

*Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556*

Abstract

We discuss a numerical method for computing the electronic scattering states for a fully two-dimensional open boundary scattering domain. The scattering states may then be used to obtain the local electron density in the vicinity of the scatterers which is necessary for a numerical study of the residual resistivity dipole and the "electron wind" force relevant to electromigration. The scattering states may also be used to calculate the local electron density of states which has recently been directly imaged by STM experiments on the surface of copper. Our numerical method is based upon the partial wave expansion of the known asymptotic form of the wave-function on the solution domain boundary. The wave-function and the normal derivative are then matched on the boundary resulting in a linear system of equations.

I. Introduction

The large volume of recent literature on the study of electromigration increasingly emphasizes the importance of the local non-uniform fields near scattering centers. It is clear that a detailed understanding of the local fields near scatterers is needed to understand phenomena in which the residual resistivity dipole (RRD) [1-2] and Friedel-oscillations [3] play an important role. Such a local field treatment is used in the application of the Kubo linear-response formalism to compute the "electron wind" force experienced by a scatterer in electromigration [4-7].

The importance of local field effects near scattering centers is most clearly and elegantly demonstrated by the recently published scanning tunnelling microscope (STM) experiments performed on the Cu(111) surface by Crommie *et al.* [8-9]. In these experiments, the local density of states (LDOS) of the two-dimensional electron gas (2DEG) on the stepped surface of Cu(111) was directly probed by an STM tip at low temperature. The images of the LDOS revealed standing-wave patterns due to electron scattering from step edges and defects.

In this paper, a numerical method is presented in which the two-dimensional electronic

scattering states are obtained by solving the effective mass Schrödinger equation over a 2D domain with an open boundary. The method uses the partial wave expansion of the known asymptotic solution and matches the wave-function and its normal derivative on the boundary. The resulting system of linear equations can be solved by the finite element method. These scattering states may then be used to compute the electron density and the LDOS inside the scattering domain. A self-consistent treatment would require an iterative solution of scattering states using a Hartree potential [10].

The physical model for the partial wave expansion is presented in section II. Section III presents the finite element formulation of the problem. An example solution for a repulsive scatterer is presented in section IV. In conclusion, we summarize our method in Section V.

II. Model

We view electron transport in the spirit of Landauer's picture for the residual resistivity dipole (RRD) [1-2] in which the incident carrier flux is identified as the fundamental driving transport quantity in a "jellium" model with a background scattering time τ which gives rise to the bulk resistivity $\rho_0 = m^*/ne^2\tau$, where n is the electron density and m^* is the electron effective mass. As is schematically depicted in Fig. 1, the incident electron flux is elastically scattered by the defect, schematically shown as the shaded spot, and is partially transmitted and partially reflected. For the metallic "jellium" model, the problem domain is assumed to be in the ballistic regime and scattering is assumed to take place within the region close to the scatterer such that $2\pi k_F l \gg 1$, where k_F is the wave vector for the electrons at the Fermi-energy, and l is the mean-free path given by $l = \hbar k_F \tau / m^*$. Outside the scattering region the domain is assumed to uniformly extend to infinity.

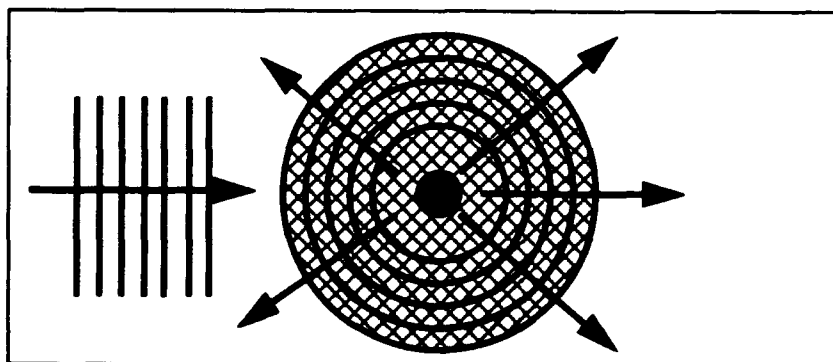


FIGURE 1: A schematic diagram of the 2-D scattering domain with an elastic scatterer at the center. The solution domain is represented by the hatched region.

Within this "jellium" model, the self-consistent electronic states can be explicitly obtained by solving the single electron effective mass Schrödinger equation,

$$-\frac{\hbar^2}{2} \nabla \cdot \left(\frac{1}{m^*} \nabla \psi_E(r, \theta) \right) + [V_R + V_H + V_{SC}] \psi_E(r, \theta) = E \psi_E(r, \theta), \quad (1)$$

where $V_R(r)$ is the potential of the scatterer, V_H is the self-consistent Hartree potential which accounts for many-body Coulombic interactions, and V_{SC} is the electrostatic potential associated with self-consistent screening. Exchange and correlation may also be included within a local density approximation [11]. The z dependence has been dropped by the assumption that only the first subband of the 2DEG is occupied, reducing the problem to circular-polar coordinates.

We assume an incident plane wave of the form $\exp(ik \cdot r)$. The asymptotic solution of this equation in the region where the scattering potential is negligible (the region outside the domain) is given by

$$\psi(r \geq R_0, \theta) = \sum_{m=-\infty}^{\infty} i^m [a J_m(kr) + b_m H_m^{(1)}(kr)] e^{im\theta}, \quad (2)$$

where we have used the Jacobi-Anger expansion of the incident wave, and J_m is the Bessel function of the first kind with a known incident amplitude a , and $H_m^{(1)}$ is the Hankel function for the outgoing scattered wave with the unknown amplitudes b_m .

Similar to the development of the quantum transmitting boundary method for quasi-1D transport [12], the 2-D transport boundary condition is developed from the orthogonality of the angular modes over the interval $\theta = (0, 2\pi)$, which are used to expand the unknown coefficients, b_m , on the boundary of the domain,

$$b_m = \frac{1}{2\pi i^m H_m^{(1)}(kR_0)} \int_0^{2\pi} e^{-im\theta} \psi(r = R_0, \theta) d\theta - a \frac{J_m(kR_0)}{H_m^{(1)}(kR_0)}. \quad (3)$$

Using this expansion for the b_m coefficients, the normal derivative of the wave-function on the boundary can be obtained,

$$\begin{aligned} \left. \frac{\partial \psi}{\partial r} \right|_{R_0} &= ak \sum_{m=-\infty}^{\infty} i^m \left(J'_m(kR_0) - J_m(kR_0) \frac{H_m^{(1)'}(kR_0)}{H_m^{(1)}(kR_0)} \right) e^{im\theta} \\ &+ k \sum_{m=-\infty}^{\infty} \frac{H_m^{(1)'}(kR_0)}{H_m^{(1)}(kR_0)} \frac{1}{2\pi} \int_0^{2\pi} e^{-im\theta} \psi(r = R_0, \theta) d\theta, \end{aligned} \quad (4)$$

and

$$\left. \frac{\partial \psi}{\partial \theta} \right|_{R_0} = \sum_{m=-\infty}^{\infty} \frac{im}{2\pi} \left(\int_0^{2\pi} e^{-im\theta} \psi(r = R_0, \theta) d\theta \right) e^{im\theta}, \quad (5)$$

where the primes on the Bessel and Hankel functions indicate derivatives with respect to kr . Equations (4) and (5) form the basis for the open 2-D scattering boundary conditions which can be incorporated into the finite element method. It is important to note that the orthogonality of the angular functions requires a circular domain boundary.

III. Numerical Method

The finite element method is used to solve the 2-D effective mass Schrödinger equation on the discretized domain schematically shown by the grided region in Fig. 1. The region

outside the problem domain is assumed to extend to infinity as described in section II. The finite element method is used to linearize the Schrödinger equation, resulting in the weak variational form

$$\hat{\mathbf{u}}^T \left(\frac{\hbar^2}{2} \int_{\Omega} \mathbf{B}^T \frac{1}{m^*} \mathbf{B} d\Omega \right) \mathbf{u} + \hat{\mathbf{u}}^T \left(\int_{\Omega} [V(r) - E] \mathbf{N}^T \cdot \mathbf{N} d\Omega \right) \mathbf{u} = \frac{\hbar^2}{2} \int_{\partial\Omega} \hat{\psi} \frac{1}{m^*} \frac{\partial \psi}{\partial r} \Big|_{\partial\Omega} \cdot \hat{\mathbf{n}}_{\partial\Omega} d\partial\Omega, \quad (6)$$

where $\hat{\mathbf{u}}$ is the vector of nodal values for the arbitrary test function $\hat{\psi}$, \mathbf{u} is the vector of unknown nodal values for the wave function ψ , \mathbf{N} is the vector of global orthonormal finite element shape functions, \mathbf{B} is the matrix of spatial derivatives of the \mathbf{N} vector of shape functions, and $\hat{\mathbf{n}}_{\partial\Omega}$ is the unit normal to the domain boundary, $\partial\Omega$. The result for the derivative of the wave-function on the domain boundary developed in section II is inserted above to determine the right hand side surface terms. The resulting linear system of equations has the form

$$\hat{\mathbf{u}}^T (\mathbf{T} + \mathbf{V} + \mathbf{C}) \mathbf{u} = \hat{\mathbf{u}}^T \mathbf{P}, \quad (7)$$

where

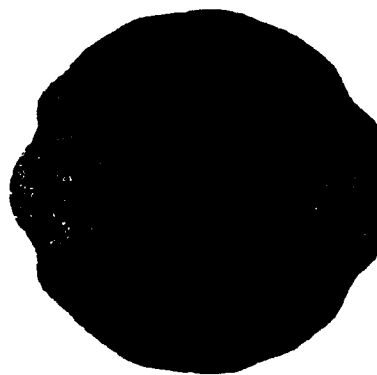
$$\begin{aligned} \mathbf{T} &= \frac{\hbar^2}{2} \int_{\Omega} \mathbf{B}^T \frac{1}{m^*} \mathbf{B} d\Omega \\ \mathbf{V} &= \int_{\Omega} [V(r) - E] \mathbf{N}^T \cdot \mathbf{N} d\Omega \\ \mathbf{C} &= \frac{\hbar^2 k R_0}{4m^* \pi} \sum_{m=-\infty}^{\infty} \frac{H_m^{(1)'}(kR_0)}{H_m^{(1)}(kR_0)} \left(\int_0^{2\pi} \hat{\psi}(r=R_0, \theta) e^{im\theta} d\theta \right) \int_0^{2\pi} e^{-im\theta} \psi(r=R_0, \theta) d\theta \\ \mathbf{P} &= a \frac{\hbar^2 k R_0}{2m^*} \sum_{m=-\infty}^{\infty} i^m \left(J_m'(kR_0) - J_m(kR_0) \frac{H_m^{(1)'}(kR_0)}{H_m^{(1)}(kR_0)} \right) \int_0^{2\pi} \hat{\psi}(r=R_0, \theta) e^{im\theta} d\theta. \end{aligned}$$

The infinite sums may be truncated to include only the more relevant low angular momentum modes with minimal loss in accuracy. The resulting linear system may be solved by standard LU-decomposition and back substitution using sparse matrix methods for the efficient use of both memory and cpu resources.

IV. A Repulsive Scatterer

Our preliminary results, neglecting electron-electron scattering effects contained in V_H and V_{SC} , are presented for the structure schematically shown in Fig. 1. The infinite repulsive scatterer is centered in the solution domain. The mesh generated for this example contains 10801 nodes connected by 21500 triangular elements. Both skyline storage and bandwidth optimization techniques were employed for an efficient computational solution of the resulting unsymmetric linear system of equations. The real part of the wave function is presented in Fig. 2. The radial and angular modes are clearly visible in the scattered wave.

FIGURE 2: The real part of the scattered wave function for an incident plane wave scattered by a repulsive central field. Both the incident plane wave and the radially scattered contributions are clearly visible. The mesh consists of 10801 nodes with 21500 triangular elements.



V. Summary

We have presented a numerical method to solve the 2-D effective mass Schrödinger equation for an open boundary scattering problem. The method uses partial wave expansion to fully specify the normal derivative of the wave-function on the boundary. The finite element method is used to discretize the Schrödinger equation. The partial wave boundary conditions are used to fully specify the problem which reduces to a linear system of equations which can be solved for the scattering states. The scattering states may then be used to compute the local density of states and the electron density inside the scattering region.

Acknowledgments

This work has been partially supported by the Office of Naval Research and the Air Force Office of Scientific Research. H.K.H. is grateful for a fellowship from the Center for Applied Mathematics of the University of Notre Dame.

References

- [1] R. Landauer, *IBM J. Res. Develop.* **1**, 223 (1957).
- [2] R. Landauer, *Phys. Rev. B* **14**, 1474 (1976).
- [3] C. Bosvieux and J. Friedel, *J. Phys. Chem. Solids* **23** 123 (1962).
- [4] R. S. Sorbello, *Phys. Rev. B* **23**, 5119 (1981).
- [5] R. S. Sorbello, *Phys. Rev. B* **31**, 798 (1985).
- [6] R. S. Sorbello and C. S. Chu, *IBM J. Res. Develop.* **32**, 58 (1988).
- [7] A. H. Verbruggen, *IBM J. Res. Develop.* **32**, 93 (1988).
- [8] M. F. Crommie, C. P. Lutz, and D. M. Eigler, *Nature* **363**, 524 (1993).
- [9] A. Zettl, *Nature* **363**, 496 (1993).
- [10] H. K. Harbury, W. Porod, C. S. Lent, *Superlat. Microstruct.* **11**, 189 (1992).
- [11] Y. Sun and G. Kirczenow, *Phys. Rev. B* **47** 4413 (1993).
- [12] C. S. Lent and D. J. Kirkner, *J. Appl. Phys.* **67**, 6353 (1990).

A Numerical Study of Resonance/Antiresonance Line Shape for Transmission in Quantum Waveguides with Resonantly-Coupled Cavities

Zhi-an Shao, Wolfgang Porod, and Craig S. Lent

Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

Abstract

We investigate the line shape of the transmission probability in quantum waveguides with resonantly-coupled cavities. The lifetime of the quasi-bound states is extracted from the asymmetrical transmission amplitude on the real-energy axis. The resonance/antiresonance feature in the vicinity of each quasi-bound state is characterized by a zero-pole pair in the complex-energy plane. We develop a generalization of the familiar symmetrical Lorentzian line shape, and discuss it in terms of Fano resonances.

I. Introduction

A common computational problem in quantum transport is to find the lifetimes of quasi-bound states from the transmission peaks. The underlying theory is that each quasi-bound state leads to a pole of the propagator (and the transmission amplitude) in the complex-energy plane. If this pole is sufficiently close to the real-energy axis, it will result in a resonance maximum of the observed transmission coefficient. A well understood problem is double-barrier resonant tunneling, where the lifetimes of the quantum well states may be extracted from the width of Lorentzian-shaped transmission peaks. A less understood problem is electronic transport in quantum waveguides with resonantly-coupled cavities [1]. It is known for these structures that the resonator states lead to resonance/antiresonance features [2, 3], but their detailed line shape has not been investigated so far. In this paper, we present a theory of the line shape for transmission in resonantly-coupled quantum waveguide, and we provide a computational method to extract the lifetimes of the corresponding quasi-bound states.

II. Poles and Zeros

For double-barrier resonant tunneling (DBRT), it is well known that the resonant transmission phenomena are related to the quasi-bound states in the quantum well region. Based on the Breit-Wigner formalism, a quasi-bound state at energy E_P and decay time $\tau = \hbar/2\Gamma$ yields a simple pole in the transmission amplitude $t(z)$ at the complex-energy $z = E_P - i\Gamma$ [4],

$$t(z) \sim \frac{1}{z - (E_P - i\Gamma)}. \quad (1)$$

If this pole is sufficiently close to the real-energy axis such that the effect of other poles can be neglected, the transmission probability, $T(E) = |t(E)|^2$, for a physical energy on the real-energy axis, E , is given by,

$$T(E) = \frac{\Gamma^2}{(E - E_0)^2 + \Gamma^2}, \quad (2)$$

which gives rise to a transmission resonance with a Lorentzian shape. It is an easy matter to extract the lifetime of the quantum well states from the width of the observed transmission peak.

An example of the transmission amplitude for DBRT is presented in Figs. 1(a) and (c), where the double-barrier structure and the transmission channel are schematically shown in the inset of Fig. 1(a). In this example, the barrier is 0.2 eV high and 3 nm thick, the separation of the barrier is 20 nm. We see the poles in the complex-energy plane are clearly visible in the contour plot of the absolute value of $t(z)$ in Fig. 1(c). The Lorentzian line shape of the transmission probability is shown in Fig. 1(a).

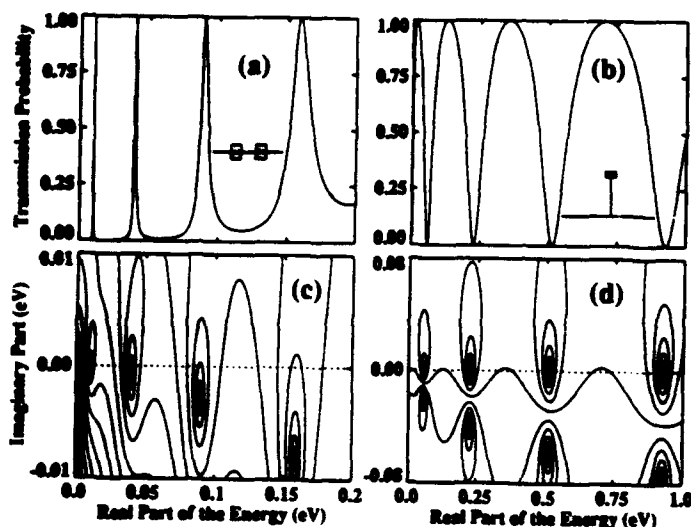


Figure 1. Comparison of the structure of the transmission amplitude in the complex energy plane for double-barrier resonant tunneling (poles) and t-stubs (zero-pole pairs). For DBRT, (a) shows the transmission probability on the real-energy axis, and (c) gives a contour plot of the absolute value of the transmission amplitude in the complex energy plane. For the t-stub structure, the corresponding plots are shown in (b) and (d), respectively.

Recently, much work has been done on transmission in resonantly-coupled quantum waveguide systems, and rich features of the transmission coefficient have been found (resonance/antiresonance) [5]. We have shown that the transmission amplitude exhibits zero-pole pairs in the complex-energy plane for this kind of the structure [6]. As a consequence, zero-pole pairs lead to asymmetrical transmission resonance/antiresonance features on the real-energy axis. As an example, we show the behavior for a t-stub structure, which is schematically shown in the inset of Fig. 1(b). It consists of a main transmission channel and a dangling wire of length $L = 10$ nm. Zero-pole pairs are clearly visible from the contour plot of the absolute value of $t(z)$ in Fig. 1(d).

III. Line Shape

Based on the zero-pole pair nature of the resonances in quantum waveguide structures, we make the following *ansatz* for the transmission amplitude in the vicinity of each quasi-

bound state,

$$t(z) \sim \frac{(z - E_0)}{z - (E_P - i\Gamma)} \quad (3)$$

Here, E_0 and $(E_P - i\Gamma)$ are the positions of the transmission zero and the pole, respectively. The lifetime of the quasi-bound state is given by $\tau = \hbar/(2\Gamma)$, as for the case of double-barrier resonant tunneling. Again, the transmission probability on the real-energy axis is given by $T(E) = |t(E)|^2$, and the proportionality constant in eqn. (3) is determined by assuming peaks with unity transmission which are known to occur in symmetrical waveguide systems [6]. A unity transmission peak at energy E_1 provides two constraints for $T(E)$, namely $T(E_1) = 1$ and $\frac{d}{dE}T|_{E_1} = 0$. It is an easy matter to show that,

$$T(E) = \left[\frac{\Gamma^2}{(E_P - E_0)^2 + \Gamma^2} \right] \left[\frac{(E - E_0)^2}{(E - E_P)^2 + \Gamma^2} \right] \quad (4)$$

The above expression gives the line shape of the transmission probability for resonantly-coupled quantum waveguides in terms of three parameters, namely the energy of the transmission zero, E_0 , the energy of the resonant state, E_P (the real part of the pole energy), and the inverse lifetime of the state, Γ (the imaginary part of the pole energy). Note that (4) produces an asymmetrical line with a resonance/antiresonance behavior. Such asymmetrical line shapes have previously been noted in atomic and molecular physics [7]. These so-called Fano resonances are known to occur when a bound state is coupled to a continuum of states, thereby leading to resonance phenomena [8].

In his original paper [7], Fano, after somewhat lengthy derivations, found that the autoionization cross section could be parameterized by $(q + \epsilon)^2/(1 + \epsilon^2)$, where ϵ is a reduced energy (it is defined as $\epsilon = (E - E_{Res})/\Gamma$, where E_{Res} is the energy of the resonant state) and q is treated as a parameter (it is a complicated expression involving matrix elements). We note that this is the same line shape as our eqn. (4) by making the following substitutions, $\epsilon = (E - E_P)/\Gamma$ and $q = (E_P - E_0)/\Gamma$. Comparing our approach to Fano's [7], we note that ϵ has a similar meaning where E_P , the real-part of the pole energy, corresponds to E_{Res} , the energy of the resonant state. For the parameter q , our approach yields a simple expression which could not have been inferred from Ref. [7]. Apparently, Fano's line shape corresponds to a zero-pole structure in the complex-energy plane, a fact which has not been noted before.

Given a certain transmission curve, we now can fit each resonance/antiresonance feature to obtain the lifetime of the corresponding quasi-bound state. Using the known energies of the transmission zero, E_0 , and transmission one, E_1 , we can find,

$$E_P = \frac{(E_0 + E_1)}{2} \pm \frac{\sqrt{(E_0 - E_1)^2 - 4\Gamma^2}}{2} \quad (5)$$

The choice of the sign in the above equation determines whether $E_P > E_0$ or $E_P < E_0$. With this, the only unknown parameter is Γ which may be used to obtain the best fit of the theoretical line shape (4) to the given transmission curve. We seek the best fit in the sense of the least mean square error.

IV. Results

We now present several examples to fit the lineshapes of resonance and antiresonance pairs. Figure 2 presents fits of the resonance/antiresonance line shapes for a family of so-called weakly coupled t-stubs [6], which are schematically shown in the insets. In Figs. 2(a), (b), and (c), the resonant stubs are separated from the main transmission channel by a tunneling barrier of length $\ell = 1.0 \text{ nm}$ and height $V_0 = 0.5 \text{ eV}$, $V_0 = 1.0 \text{ eV}$, and $V_0 = 2.0 \text{ eV}$, respectively. In each case, we show 3 quasi-bound states which lead to zero-one features in the transmission probability, and which are labeled in the plots. Figures 2(α), (β), and (γ), show the fitted line shape for the resonance numbered 3 of cases (a), (b), and (c), respectively. The fit is shown by the dotted line, and the curve to be fitted by the solid line.

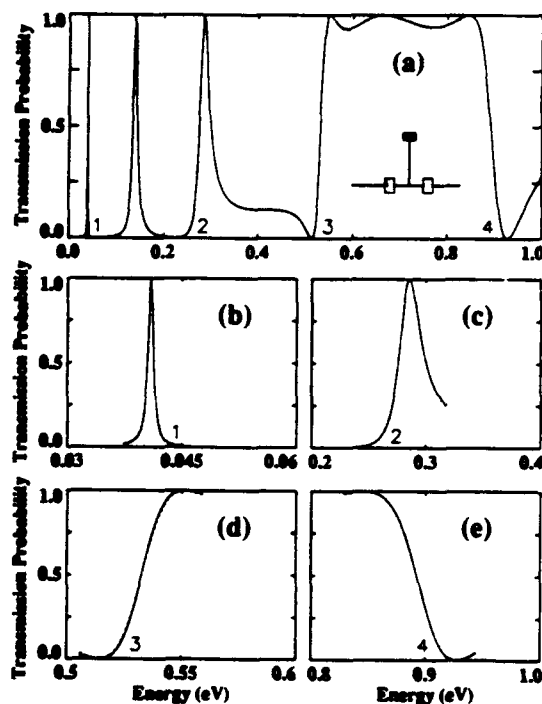


Figure 3. Example of fits using the line shape (4) for a t-stub with a double barrier structure on the transmission channel shown in the inset to (a). The fits of resonances 1 - 4 are shown in (b) - (e), respectively.

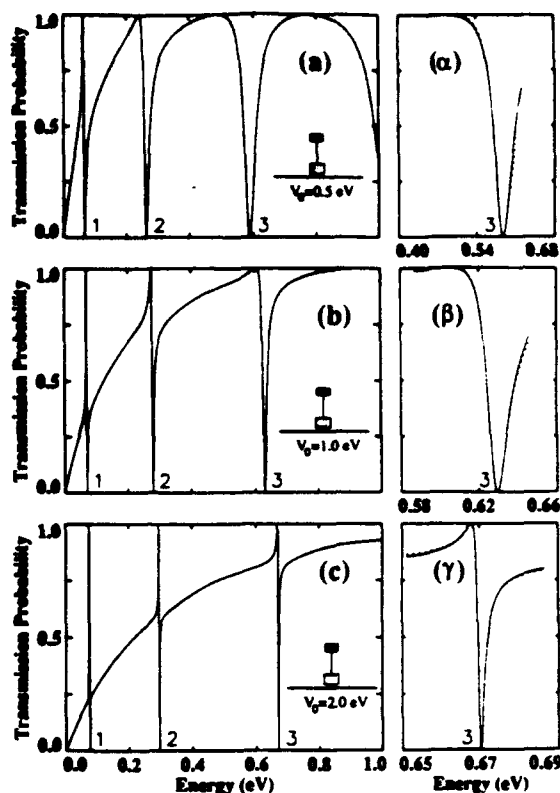


Figure 2. Examples of fits using the line shape (4) for the weakly-coupled t-stub structures shown in the insets. The fits to the third resonant state of (a), (b), and (c) are given in (α), (β), and (γ), respectively. The fits are shown by the dotted line, and the curve to be fitted by the solid line.

In figure 3, we present another example which shows transmission for a t-stub in addition to double-barrier resonant tunneling on the main transmission channel. A schematic drawing of this waveguide structure is displayed in the inset, and the two tunneling barriers have a thickness of 1 nm , height of 0.5 eV , and separation of 4 nm . Figure 3(a) shows the transmission probability on the real-energy axis. Figures 3(b) - (e) display the fits according to our *ansatz*, eqn. (3), for resonances 1 - 4, respectively. Again, the fit is shown by the dotted line, and the curve to be fitted by the solid line. It appears that the zero-pole character of each quasi-bound state, leads to extremely good fits of the transmission probability in the vicinity of each resonance.

In general, the locations of the poles and the zeros on the real-energy axis are not the same, i.e. $E_P \neq E_0$. It is this fact that gives rise to the asymmetric line shape, eqn. (4). Note that from eqn. (5), the position of the pole, E_P , is always between the positions of the transmission zero and one, E_0 and E_1 . If the pole and the zero occur at the same real energy, i.e. $E_P = E_0$, then eqn. (4) yields a symmetric line shape,

$$T|_{E_P=E_0} = \frac{(E - E_P)^2}{(E - E_P)^2 + \Gamma^2} \quad (6)$$

The above expression describes a Lorentzian-shaped reflection line.

In recent work [3], Price has pointed out that a resonant quasi-bound state can give rise to either Lorentzian-shaped transmission or reflection peaks, and he terms these peaks resonances of the first and the second kind, respectively. We see that the reflection peaks in general will not have a Lorentzian shape, and that Price's resonances of the second kind are recovered when $E_P = E_0$.

In summary, we have investigated the detailed line shape of the transmission probability in quantum waveguides with resonantly-coupled cavities. The resonance/antiresonance features in the vicinity of each quasi-bound states can be characterized by a zero-pole pair in the complex-energy plane. We have found a generalization of the familiar symmetrical Lorentzian resonance peaks. Using several examples, we have demonstrated the utility of our line shape (4) to extract the lifetime of the quasi-bound state by a fit to the data. We also discussed the asymmetrical line shapes in the context of Fano resonances.

Acknowledgement: This work was supported in part by AFOSR, ARPA, and ONR.

References

- [1] *Nanostructure Physics and Fabrication*, ed. by M. A. Reed and W. P. Kirk (Academic Press, Boston, 1989).
- [2] C. S. Lent, *Computational Electronics: Semiconductor Transport and Device Simulation*, ed. by K. Hess, J. P. Leburton, and U. Ravaioli, p. 259 (1991).
- [3] P. J. Price, *Appl. Phys. Lett.* **62**, 289 (1993), and preprint.
- [4] T. B. Bahder, C. A. Morrison and J. D. Bruno, *Appl. Phys. Lett.* **51**, 1089 (1987).
- [5] F. Sols, M. Macucci, U. Ravaioli, K. Hess, *Appl. Phys. Lett.* **54**, 350 (1989) and *J. Appl. Phys.* **66**, 3892 (1989).
- [6] W. Porod, Z. Shao and C. S. Lent, *Appl. Phys. Lett.* **61**, 1350 (1992).
- [7] U. Fano, *Phys. Rev.* **124**, 1866 (1961).
- [8] E. Tekman and P. F. Bagwell, *Phys. Rev. B* **48** (July 15 - 1993).

Effects of Optical and Acoustic Phonon Scattering on Hot Electron Transport in Quantum Wires

N. Mori, H. Momose, S. Miyahara and C. Hamaguchi

Department of Electronic Engineering, Osaka University
2-1, Yamada-oka, Suita City, Osaka 565, Japan

Abstract

Transport responses of hot electrons in quantum wires at high temperature are calculated including both optical and acoustic phonon scatterings to study how the average electron energy and the drift velocity depend on the applied electric field. The average electron energy is found to increase monotonically with the applied electric field showing no anomalous carrier cooling. The calculated results of drift velocity at lower electric fields agree well with the results obtained by solving linearized Boltzmann transport equation. The contribution of acoustic phonon scattering to the transport response is also discussed.

I Introduction

In recent years, mesoscopic structures, such as quantum wires and dots, have been designed to investigate new phenomena and to explore high speed devices. In quantum wires, momentum space is limited to a single dimension, and the carrier density of states (DOS) shows some singularities resulting in changes in the nature of particle collisions. It is pointed out that at a moderate electric field anomalous carrier cooling of electrons in quantum wires enhances the transport response due to the conversion of thermal energy into drift motion [1]. In the present work, we are interested in obtaining transport responses of hot electrons in quantum wires at high temperature including both optical and acoustic phonon scatterings to study how the average electron energy and the drift velocity depend on the applied electric field. We adopt the Rees's iterative method [2] to evaluate hot electron distribution function, since the method provides a consistent treatment of the divergence of DOS of electrons and the scattering probability [3].

II Model

We employ a simple model for a quantum wire in which a two-dimensional electron gas in xy -plane is confined by narrow gates or split gates, and electrons are free along only the x direction. We assume that the wave function associated with quantized z motion is expressed by the Fang-Howard variational function ($\xi_0(z) = (b^3/2)^{1/2} z \exp(-\frac{1}{2}bz)$) and the confinement in the y direction is characterized by a parabolic potential of frequency Ω ($V(y) = \frac{1}{2}m\Omega^2 y^2$). We consider only the ground subband, i.e. the strong confinement case. Furthermore, number of electrons is assumed to be so small that their distribution is given by the Boltzmann distribution function in thermal equilibrium.

III Method

We evaluate the electron distribution function by using the iterative method proposed by Rees [2] accounting the Fermi's golden rule to calculate the scattering probability. In this study, we consider bulk longitudinal optical phonon scattering and acoustic deformation potential scattering, while other scattering processes are ignored for simplicity. The distribution function $f(k)$ is calculated by the iterative process as follows:

$$f_{n+1}(k) = \Gamma \int_0^\infty g_n(k - eEt) e^{-\Gamma t} dt, \quad (1)$$

$$g_n(k') = \sum_k f_n(k) \frac{S(k, k')}{\Gamma}, \quad (2)$$

where $S(k, k')$ is the scattering probability per unit time from k to k' and Γ is the total scattering rate including self-scattering process which is positive constant and determines the convergence of the iteration. $S(k, k')$, which includes self-scattering process, is evaluated from *real* scattering probability $W(k, k')$ as follows:

$$S(k, k') = W(k, k') + S(k) \delta_{k, k'} \quad (3)$$

with $S(k)$ being the self-scattering rate. $W(k, k')$ is given by the sum of optical and acoustic phonon scattering probabilities:

$$W(k, k') = \sum_{\eta=\pm 1} 4\pi\alpha\omega_0 \left(\frac{\omega_0}{k_0}\right) (N_0 + \frac{1}{2} + \frac{1}{2}\eta) G(k' - k) \delta(\epsilon_{k'} + \eta\omega_0 - \epsilon_k) \\ + 8\pi\beta\omega_0 \left(\frac{\omega_0}{k_0}\right) \left(\frac{kT}{\omega_0}\right) f_y f_z \delta(\epsilon_{k'} - \epsilon_k) \quad (4)$$

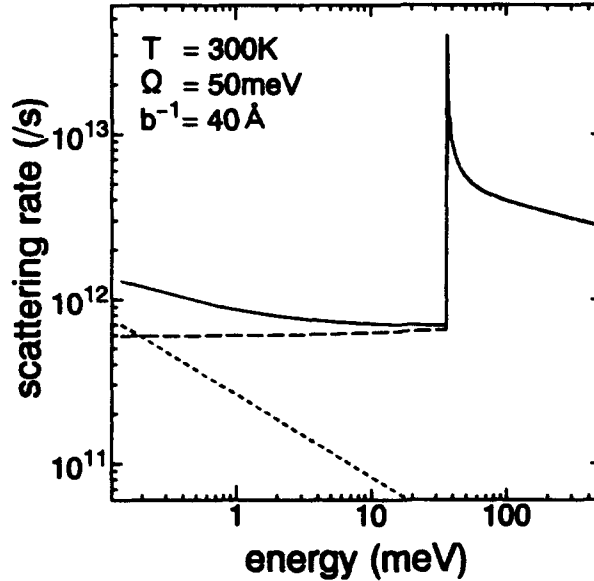


Figure 1: Total scattering rates in a GaAs quantum wire (solid line). $1/b = 40 \text{ \AA}$, $\Omega = 50 \text{ meV}$ and $T = 300 \text{ K}$. Dashed line : rates for optical phonon scattering. Dotted line : rates for acoustic phonon scattering.

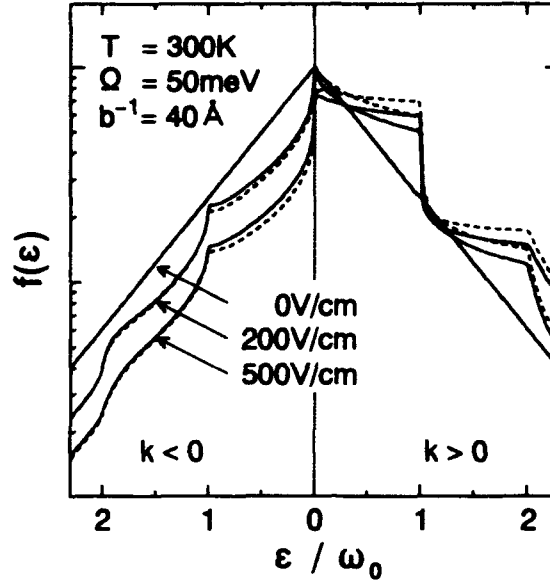


Figure 2: Distribution functions $f(k)$ as a function of electron energy $\varepsilon_k = k^2/2m$ with applied electric fields (0, 200 and 500 V/cm) switched on for 100 ps. $1/b = 40 \text{ \AA}$, $\Omega = 50 \text{ meV}$ and $T = 300 \text{ K}$. Solid lines represent $f(k)$ including both optical and acoustic phonon scatterings, and dotted lines are $f(k)$ without acoustic phonon scattering.

with

$$G(q_x) = \int_{-\infty}^{+\infty} dq_y \frac{1}{2Q} e^{-(q_y/K)^2} F(Q), \quad F(Q) = \iint |\xi_0(z_1)|^2 |\xi_0(z_2)|^2 e^{-Q|z_1-z_2|} dz_1 dz_2, \quad (5)$$

where α is the Fröhlich coupling constant, ω_0 is the optical phonon energy, N_0 is the occupation number of optical phonons, $\beta = D^2 m k_0 / 2 \rho s^2$, D is the deformation potential, ρ is the density of the material, s is the sound velocity, $f_y = (K/k_0)/(2\sqrt{\pi})$, $f_z = (3/16)(b/k_0)$, $K = (2m\Omega)^{1/2}$, $k_0 = (2m\omega_0)^{1/2}$ and $Q = (q_x^2 + q_y^2)^{1/2}$. Note that non-elastic effects of acoustic phonon scattering are ignored in the present calculation.

Calculated scattering rates of a GaAs quantum wire are shown in Fig. 1. Scattering rates by acoustic phonons are larger than those of optical phonons for low energy electrons because of the singularity of DOS at the bottom of subband.

IV Results

Figure 2 shows calculated results of distribution functions with applied electric fields (0, 200 and 500 V/cm) switched on for 100 ps. In the case of zero applied electric field, the distribution function coincides with the Boltzmann distribution function. This clearly indicates that the iterative method is valid for evaluating transport responses of one-dimensional electron gases.

In Fig. 3 we show the calculated average electron energy ε_{ave} in a quantum wire of GaAs as a function of applied electric field E . We find that ε_{ave} increases monotonically with E and an anomalous carrier cooling does not occur. The calculated result of drift

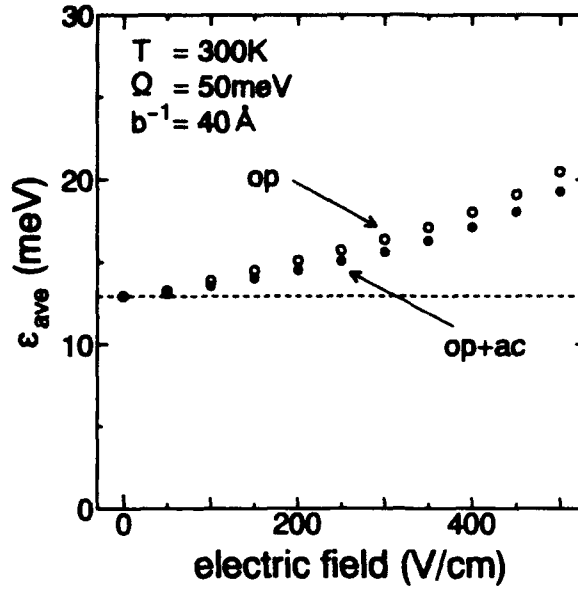


Figure 3: Average electron energy ϵ_{ave} in a GaAs quantum wire as a function of applied electric fields. Thermal kinetic energy ($\frac{1}{2}kT$) is represented by dotted line. $1/b = 40 \text{ \AA}$, $\Omega = 50 \text{ meV}$ and $T = 300 \text{ K}$. Solid circles indicate ϵ_{ave} including both optical and acoustic phonon scatterings, and open circles are ϵ_{ave} without acoustic phonon scatterings.

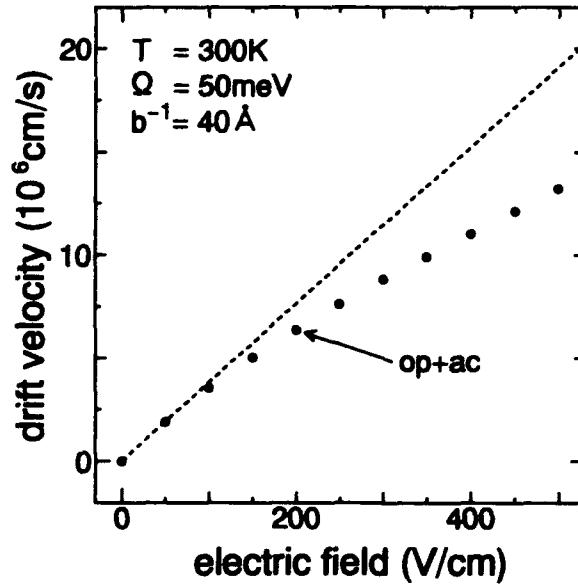


Figure 4: Electron drift velocity in a GaAs quantum wire as a function of applied electric fields (solid circles), compared with the results obtained by solving linearized Boltzmann transport equation (dotted line). Both optical and acoustic phonon scatterings are considered. $1/b = 40 \text{ \AA}$, $\Omega = 50 \text{ meV}$ and $T = 300 \text{ K}$.

velocity is also shown in Fig. 4. The drift velocity at lower electric fields agrees well with the results obtained by solving linearized Boltzmann transport equation. At higher electric fields, the drift velocity saturates as the electron temperature increases. Although electron energy and drift velocity are influenced only a little by acoustic phonon scattering in the case of $1/b = 40 \text{ \AA}$ and $\Omega = 50 \text{ meV}$, it is not too small to be neglected.

Figure 5 shows a contribution of acoustic phonon scattering to the drift velocity. The contribution, η_{ac} , is defined as follows:

$$\eta_{ac} = 1 - \frac{v_{op+ac}}{v_{op}} \quad (6)$$

where v_{op+ac} is the drift velocity calculated by including both optical and acoustic phonon

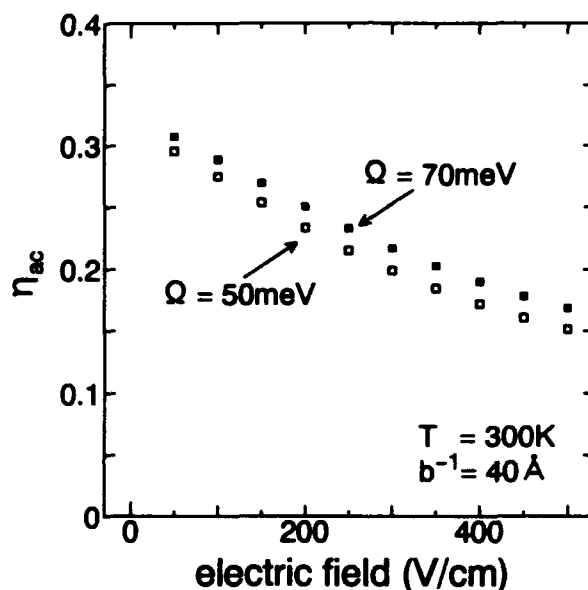


Figure 5: Contribution of acoustic phonon scattering to the drift velocity for $\Omega = 50$ and 70 meV . The contribution, η_{ac} , is defined in the text. $1/b = 40 \text{ \AA}$ and $T = 300 \text{ K}$.

scatterings, and v_{op} is that without acoustic phonon scattering. The contribution of acoustic phonon scattering becomes larger in the case of low applied electric field E or strong confinement Ω .

References

- [1] J.P. Leburton and D. Jovanovic, *Semicond. Sci. Technol.* **7** (1992) B202.
- [2] H.D. Rees, *J. Phys. Chem. Solids* **30** (1969) 643.
- [3] N. Mori and C. Hamaguchi, in *Extended Abstracts of the Second International Symposium on New Phenomena in Mesoscopic Structures*, Maui, Dec. 7-11, 1992, pp. 28-31.